PHD

Facial emotion expression, recognition and production of varying intensity in the typical population and on the autism spectrum.

Wingenbach, Tanja

Award date:
2016

Awarding institution:
University of Bath

Link to publication

Facial emotion expression, recognition and production of varying intensity in the typical population and on the autism spectrum

Tanja S. H. Wingenbach

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Psychology

February 2016

Table of Contents

TABLE OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to thank everyone who has accompanied me during my time as a PhD Candidate in any way. Special thanks goes to Mark Pellowe who always lent me an ear and proof-read this dissertation. Gratitude goes to my supervisors Mark Brosnan and Chris Ashwin, especially Chris who continuously pushes me to achieve the best possible outcome. I am also very grateful to my mother (Helga Wingenbach) and father (Michael Schneider) for always being there when needed. I further want to acknowledge the precious conversations I had with Ian Fairholm and Oliva Bolt. I am very lucky to have such great people in my life – thank you all.

Great gratitude goes to the various funding bodies that supported me and my research project, the FAZIT-Stiftung, the German Academic Exchange Service, and the University of Bath. I further want to thank all individuals who volunteered to participate in my studies and the Undergraduate students who assisted in data collection during various parts of this research project (Jack Barnett-Leveson, Alicia Cork, and Lauren Huff). I am also thankful to the developers of the video stimuli that were used within the current research project for giving me permission to use the videos. Last but not least I want to thank my research collaborators, especially Monique Pfaltz, for their support and passing on their knowledge to me.

PUBLICATIONS

Parts of the work within this dissertation have been published, as conference posters, talks, or as a journal publication.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV): A Set of Videos Expressing Low, Intermediate, and High Intensity Emotions. *PLoS ONE 11(1).* e0147112. doi:10.1371/journal.pone.0147112

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (July 2015). *Reduced accuracy in recognising basic and complex emotions of varying intensities in high-functioning Autism Spectrum Disorders*. Poster presented at International Society for Research on Emotions (ISRE), Geneva, CH.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (May 2015). *Facial Emotion Recognition from Videos with Varying Intensity in Autism*. Poster presented at International Meeting for Autism Research (IMFAR), Salt Lake City, UT, USA.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2015, May). *Recognition of facial emotional expressions of varying intensities in high-functioning autism*. Presented at the University of Bath Psychology PhD Conference, Bath, UK.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2014, May). *Dynamic assessment of facial emotion recognition at varying intensities*. Presented at the University of Bath Psychology PhD Conference, Bath, UK.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2014, March). *Development and validation of a more ecologically valid facial emotion recognition paradigm*. Presented at the University of Zurich, Clinic for Psychiatry and Psychotherapy research colloquium, Zurich, CH.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2013, December). *Facial emotion recognition – Different methodological approaches and applications to autism – Facial emotion recognition using natural face videos*. Presented at the Cognitive Affective Social Technology Laboratories research colloquium, Bath, UK.

Wingenbach, T. S. H., Brosnan, M., & Ashwin, C. (August 2013). *Developing an ecologically valid facial emotion recognition experiment*. Poster presented at International Society for Research on Emotions (ISRE), Berkeley, CA, USA.

ABSTRACT

The current research project aimed to investigate facial emotion processing from especially developed and validated video stimuli of facial emotional expressions including varying levels of intensity. Therefore, videos were developed showing real people expressing emotions in real time (anger, disgust, fear, sadness, surprise, happiness, contempt, embarrassment, contempt, and neutral) at different expression intensity levels (low, intermediate, high) called the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV). The ADFES-BIV was validated on all its emotion and intensity categories. Sex differences in facial emotion recognition were investigated and a female advantage in facial emotion recognition was found compared to males. This demonstrates that the ADFES-BIV is suitable for investigating group comparisons in facial emotion recognition in the general population. Facial emotion recognition from the ADFES-BIV was further investigated in a sample of individuals that is characterised by deficits in social functioning; individuals with an Autism Spectrum Disorder (ASD). A deficit in facial emotion recognition was found in ASD compared to controls and error analysis revealed emotion-specific deficits in detecting emotional content from faces (sensitivity) next to deficits in differentiating between emotions from faces (specificity). The ADFES-BIV was combined with face electromyogram (EMG) to investigate facial mimicry and the effects of proprioceptive feedback (from explicit imitation and blocked facial mimicry) on facial emotion recognition. Based on the reverse simulation model it was predicted that facial mimicry would be an active component of the facial emotion recognition process. Experimental manipulations of face movements did not reveal an advantage of facial mimicry compared to the blocked facial mimicry condition. Whereas no support was found for the reverse simulation model, enhanced proprioceptive feedback can facilitate or hinder recognition of facial emotions in line with embodied cognition accounts.

LIST OF ABBREVIATIONS

ADFES	Amsterdam dynamic facial expression set
ADFES-BIV	Amsterdam dynamic facial expression set – Bath intensity variations
APA	American Psychiatric Association
ASD	Autism Spectrum Disorder
AU	Facial action unit
BM	Blocked facial mimicry with a facial emotion recognition task
EMG	Electromyogram
FACS	Facial action coding system
FER	Facial emotion recognition (no face movement manipulation)
GLMM	Generalised linear mixed modelling
IM	Explicit imitation with a facial emotion recognition task
IQR	Interquartile range
p value	Item difficulty index / probability value
SM	Spontaneous facial mimicry (without facial emotion recognition)

DEFINITION OF TERMS

apex	the maximum intensity of an emotional facial expression
arousal	dimension ranging from low (calm) to high (excited)
basic emotions	universal, evolved, and adaptive emotions (anger, disgust, fear, sadness, surprise, happiness)
blocked facial mimicry	face movement manipulation by holding a pen with the lips
complex emotions	non-basic emotions (e.g. embarrassment, pride, contempt)
facial emotion recognition	ability to detect and correctly identify emotions from faces
facial mimicry	spontaneous, automatic, and unconscious subtle activation of facial features in response to observed expressions of facial emotion
explicit imitation	instructed exact imitation of observed expressions of facial emotion
negative emotions	emotions of negative valence
offset	the change from an emotional expression back to a neutral expression
onset	the change from a neutral facial expression to an emotional facial expression
positive emotions	emotions of positive valence
sensitivity	ability to detect emotions in faces
specificity	ability to differentiate between emotions in faces
valence	dimension ranging from unpleasant (negative) to pleasant (positive)

LITERATURE REVIEWS

CHAPTER 1—Emotions and facial emotion recognition

Humans are very social and in social interactions our attention is often drawn to faces (Theeuwes & Van der Stigchel, 2006). This can be explained by the multitude of information faces convey about others, which includes such invariant aspects as sex (Bruce et al., 1993), race (Freeman et al., 2015), and age (Rhodes, 2009). Faces also show dynamic features, facial expressions. One role of facial expressions is to confer emotional information. Since facial expressions can be interpreted by an observer, emotional facial expressions can serve a communicative role (Fridlund, 1994). That facial expressions are used as a means of communication is supported by research showing that when an individual is alone facial expressions either do not occur, or occur only in a very subtle form (e.g. Fridlund, 1991). Facial expressions as a means of communication also carry inter-personal information, enabling promotion of bonding as well as the development and regulation of interpersonal relationships (Ekman, 2005). Facial expressions of emotion can be used to regulate the environment by indicating people's intentions and actions (Horstmann, 2003). For example, an angry facial expression communicates dominance (Keating et al., 1981) and rejection (Heerdink, van Kleef, Homan, & Fischer, 2015), which facilitates avoidance, whereas a fearful face conveys submissiveness and facilitates approaching (Marsh, Ambady, & Kleck, 2005). However, when facial emotional expressions are used in a more functional way for social regulation, expressions do not necessarily have to accurately reflect the current emotional state (Fridlund, 1994). Either way, emotional information from the face is used in social interactions to understand the feelings and intentions of others. As such, being able to interpret these expressions contributes to social functioning. It is not surprising that people vary in their ability to express and interpret facial expressions of emotions. The ability to detect the emotional cues in the face and interpret them correctly is called *facial emotion recognition* and has attracted a vast amount of research in typically developed individuals as well as in clinical populations such as the Autism-Spectrum-Disorder (ASD), as these individuals have deficits in non-verbal social communication. Emotional states can also be expressed vocally, by gestures, and body postures (Boucher & Carlson, 1980; Ruffman, Henry, Livingstone, & Phillips, 2008), but given the importance of facial emotional expressions for social interactions (K Scherer & Scherer, 2011) and for conveying crucial information about others and ourselves (Marsh, Kozak, & Ambady, 2007), the face has gained particular attention within emotion research including ongoing research. Due to the importance of the face in social interaction, the current research project focused on emotion recognition from the face only, not from other expression modalities. The aim of this chapter is to define the terminology used throughout this dissertation relating to emotion and provide a theoretical background on the early emotion theory that has influenced contemporary research on emotion leading up to the current research project.

Theories of emotion.

Charles Darwin was one of the most influential people for the development of theories of emotion and research on emotions. Darwin (1872/1965) studied emotions and the expression of emotions across species with a view to demonstrate that human expressions of emotions have a genetic basis, are evolved and adaptive. That is, the emotions that are present in humans have evolved through processes of adaptation. Thereby, emotions are considered responses to a stimulus or situation that have been functional in the course of evolution in that they secured survival (phylogeny). According to Darwin (1872/1965), expressions of emotions can be seen in humans and other animals giving them a universal character. The function of expressions of emotion can be derived from this universality. Only those expressions that proved functional in evolution by securing survival remained (this process is called natural selection) and these expressions of emotion are therefore considered to have protective functions. As such, expressions of emotions allow us to process life challenges we experience. As a consequence of this view of emotions, Darwin (1872/1965) proposed that expressions of emotion are directly linked to the feeling of an emotion. The experience of an emotion is based on the changes that occur within the body once an emotion is elicited, e.g. the face turning red during anger. These changes in the body are visible to an observer and provide information for the observer. For emotional expressions to send an interpretable signal and serve as a means of communication, the emotion needs to be expressed a certain way for it to be clearly attributable to a specific emotion. This assumption was derived from Darwin's work and led to an understanding of individual emotions as being distinct and entailing differentiable patterns.

Several theories have been postulated that consider emotions as distinct entities and proposed some emotions as being biologically hard-wired as suggested by Darwin (e.g. Ekman, Friesen, & Ellsworth, 1982; Izard, 1977; Plutchik, 1980; Tomkins, 1984). An emotion researcher of particular importance in regards to studying distinct emotions is Paul Ekman. His theory of basic emotions (Ekman, 1992a, 1992b) has proven very influential in the field of emotion research and also gained popularity in the general population through the media. Ekman calls the evolved hard-wired emotions basic emotions and suggests the existence of seven basic emotions: anger, fear, sadness, disgust, surprise, happiness, and contempt (Ekman & Cordaro, 2011), although contempt is the most contentious one and was at first not considered a basic emotion by Ekman. The term basic emotions is from now on used throughout this dissertation referring to evolved distinct emotions. It has to be mentioned that the theories of basic emotions all differ in the amount of basic emotions they propose and even the number of basic emotions proposed by Ekman varied over the years. Ortony and Turner (1990) reported that the amount of proposed emotions varies between two and 18. It is these variations in reports that has led Ortony and Turner (1990) to deny the existence of basic emotions completely, although they raise some more arguments against the basic emotions theory in their review article. It has to be noted though that substantial overlap exists in regards to which emotions are proposed by the varying basic emotion theorists. The overlap becomes even more apparent when considering emotions as families as suggested by Ekman (1993), e.g. rage and frustration are both part of the anger family but differ in their intensity. The emotions that are most frequently reported as basic emotions are anger, fear, sadness, and happiness (Ortony & Turner, 1990). Despite the variations in the proposed number of basic emotions, there seems to be consensus about the just mentioned four emotions.

Since basic emotions are supposed to present themselves in a distinct manner, the existence of basic emotions can be tested by investigating differentiable patterns. What is meant by differentiable patterns for individual emotions is exemplified here on the example of fear. When an emotion is elicited, changes occur in the body including brain activity, facial expressions, vocal expressions, body physiology, subjective experience, and actions, with varying intensity. When fear is elicited, then changes occur in brain activity in the amygdala (a region in the brain associated with emotion processing in general, but specifically with fear). The associated facial expression includes wide opened eyes, eyebrows pulled upwards and drawn together and the corners of the mouth pulled outwards. Changes in physiology manifest in the face turning pale and a sweat response. The subjective experience of fear leads to the direction of attention towards survival. According to Tomkins (1962), specific response patterns are elicited automatically by certain events. In the example of fear, the event would be a threat or danger in the close environment or the memory thereof. If the outlined pattern is indicative of fear and fear is a basic emotion, the pattern should not occur with any other emotion.

Since the definition of basic emotions claims that there are distinct patterns for each emotion that are universal, studies were conducted to explore emotion-specific patterns. One approach to investigate the existence of basic emotions is to test their recognisability. Due to the assumed universality of these emotions they should be recognised by everyone regardless of their culture or race. The universality of the basic emotions including their facial expressions would mean that there are prototypes for these expressions. Driven by the argument that exposure to mass media leads to the universality in emotions in literate countries rather than a basis in evolution, Ekman and Friesen (1971) conducted a study on emotion judgement of the six basic emotions proposed by Ekman (without contempt) in New Guinea with a sample from a preliterate culture group, not influenced by media and minimal contact to Westerners. Participants were read a story by a translator and had to choose one out of three presented photographs of facial emotional expression they thought was portrayed by the story. Given the assumption that each basic emotion has its distinct pattern how it presents itself in the body, the facial expressions should be clearly attributable to the emotion told by the story. A forced multiple-choice answer format was applied by presenting several photographs to the participant who had to select the image that best represented the story. This answer format is still widely used today. The results showed high accuracy rates, that is, the correct faces were matched to the stories most of the time for the proposed six basic emotions. However, it should be noted that attributions of the surprise face to a fear story happened to a substantial degree when a surprise face was presented as one of the answer choices. Confusions of facial expressions did not occur to such a degree when fear was presented with other facial expressions of emotion.

Nonetheless, the high confusion rate should not be interpreted as a lack of distinction between fear and surprise and that surprise therefore cannot constitute a basic emotion. This is because there might be cultural differences in regards to which situation elicits which emotion, as argued by Ekman and Friesen (1971). This possibility is important to consider, since the participants were asked to match a face to a story and not to give a name to a displayed facial expression of emotion. The attribution of a label to an emotional face is to be differentiated from an attribution of a facial emotion to a story, as the latter reflects which expression is shown by a certain eliciting event. The question is whether it is required for a basic emotion that the same label is attributed to an emotional facial expression universally or that the situation eliciting that specific emotion is universal. The latter is more prone to individual differences, the habitual environment, and cultural norms and it is therefore possible that the participants from New Guinea would react with surprise to some of the events told by the fear-stories. That does not mean that fear is not universally recognised. It is also possible that surprise was attributed to fear, as surprise often precedes fear (Ekman & Friesen, 1975); a sudden appearance of a threat elicits surprise which then turns into fear. This could explain the response of surprise to some of the fear-stories. Because of the high recognition rates for most of the conditions and the alternative explanation of the confusion of fear with surprise, Ekman and Friesen (1971) interpreted the results from their study as evidence for universality of the six tested facial expressions and as support for the basic emotion theory.

However, since Ekman and Friesen (1971) only presented participants with two non-target expressions alongside the target expression, the non-target expressions were not always emotions that were most similar to the target expression. This could have led to some correct answers being facilitated. However, Ekman and Friesen (1971) presented at least one emotion that was likely to be confused next to the target emotion, attempting to account for biased accuracy rates based on the answer choices provided. Confusions of facial expressions of certain basic emotions with one another are common and particularly likely for emotional facial expressions that share facial features (e.g. wide open eyes in fear and surprise). As long as confusion rates are lower than recognition rates and recognition rates substantially higher than the chance level of responding, confusions do not oppose the basic emotion theory (Ekman, 1994). Nevertheless, the study provides strong evidence for the six basic emotions despite some limitations. The results are also in line with earlier reports from similar studies across cultures where participants had to attribute emotion labels to emotional faces depicting the six basic emotions and high agreement rates were retrieved (e.g. Ekman & Friesen, 1969; Ekman, Sorenson, & Friesen, 1969; Izard, 1968; Izard, 1969). Furthermore, Ekman (1971) videotaped participants from New Guinea whilst displaying the six basic emotions each facially and presented these videos to Americans, who were asked to label the displayed emotions. The facial emotional expressions displayed by the New Guinean participants were found to be recognisable by the American participants (except for confusions of fear and surprise). This result serves as further indicator of universality of the facial expressions of the six basic emotions.

In the 'Facial Action Coding System' (FACS; Ekman & Friesen, 1978; new edition: Ekman, Friesen, & Hager, 2002) all possible facial movements that are the result of face muscle movements are catalogued and are called 'Action Units' (AU). Ekman, Friesen, and Hager (2002) provided suggestions for AU combinations that align to basic emotional expressions. For example, the activations of AU 9 (nose wrinkle), AU 10 (upper lip raise), and AU25 (lips parted) express disgust. It has to be noted that single AUs are independent from emotions and presumably this is the reason why the FACS is used by emotion researchers regardless with which emotion theory they affiliate themselves. Ideally, there would be a definite list containing the six basic emotions and the AUs that underlie their facial expression prototypes that is agreed upon. Since no such unified list exists, there are slight variations in what individual researcher consider as core AUs for specific emotional expressions.

Further evidence for the existence of at least six basic emotions comes from investigations on a physiological level instead of on a cognitive level as is the case when matching faces to stories. Distinct patterns in physiology on basis of heart rate, temperature, and electrodermal activity were found for the emotions anger, fear, sadness, happiness, surprise, and disgust (see Ekman, Levenson, & Friesen, 1983; Levenson, Carstensen, Friesen, & Ekman, 1991; Levenson, Ekman, & Friesen, 1990). The varying physiological patterns can be linked to functions of emotions on a behavioural level (as proposed by Darwin). In anger, a preparation for fighting occurs by increasing the blood flow to the hands (Levenson et al., 1990). In fear, the blood flow to large skeletal muscles increases which prepares for a flight reaction (Levenson et al., 1990). Disgust will lead to a rejection of the stimulus by restricting airflow to olfactory receptors and triggering a gag reflex (Koerner & Antony, 2010). Sadness results in a loss of muscle tone (Ekman & Friesen, 2003), slowing us down, allowing to focus on the issue that induced the sadness (Wolpert, 2008). During happiness, the available energy is increased by releasing the respective transmitters (Uvnäs-Moberg, 1998). During surprise, air is quickly inhaled and the ability to react quickly increased (Ekman & Friesen, 1975), as it interrupts ongoing processes (Tomkins, 1962). Due to the advanced technology now available, brain imaging techniques can be applied and allows for investigation of the existence of basic emotions also at a level of neural substrates. Vytal and Hamann (2010) conducted a neuroimaging meta-analysis to investigate distinct activity patterns in the brain for anger, disgust, fear, happiness, and sadness and found indeed distinct patterns of neural correlates for each of the emotions investigated. Distinct patterns for each of the six basic emotions were also found by a study asking participants to colour in the body parts they thought are affected by either increase or decrease in sensations when feeling the six basic emotions and the obtained results were in line with associated physiological changes as outlined above (see Nummenmaa, Glerean, Hari, & Hietanen, 2014). The evidence presented here on distinct patterns and functions that can be linked to specific emotions serve as support for the existence of basic emotions.

The presented evidence leads to assume that there are six basic emotions (anger, disgust, fear, sadness, surprise, and happiness), but human experience dictates that there exist more emotions than those six emotions. This fact does not oppose the basic emotion theory, as the emotions that do not fit the definition of basic emotions are assumed to be species-specific and to vary between cultures. Such non-basic emotions are called *complex emotions*. According to some researchers (Ekman & Friesen, 1975; Oatley & Johnson-laird, 1987; Plutchik, 1980), complex emotions are blends of basic emotions. Blends could arise as a result of two basic emotions being elicited, for example, contempt as a complex emotional expression could arise from a mixture of disgust and anger. Blends could also be the result of an attempt to neutralise a facial expression of a certain emotion and to replace it by an acceptable one (called masking) based on cultural display rules. The concept of emotion blends is yet to receive much attention from the emotion research community. Instead, complex emotions are often defined as emotions that are also distinct but include a greater cognitive component than basic emotions. Some complex emotions are commonly known as *self-conscious emotions* (e.g. embarrassment), indicating the necessity of self-awareness and the involvement of more complex cognitive processes, such as self-evaluation and assumptions about how others perceive oneself (Tracy & Robins, 2007). It is suggested that the function of complex emotions is mainly to regulate social behaviour and, as such, are also called *social emotions* or *moral emotions* (Adolphs, 2002). Within this dissertation the term complex emotion is used when referring to non-basic emotions. Therefore, throughout this dissertation a classification into basic and complex emotions for the emotions investigated within this research project is used with anger, disgust, fear, happiness, sadness, and surprise as basic emotions, and embarrassment, contempt, and pride as complex emotions; even though there is a substantial debate about whether contempt is a complex emotion or a basic one as proposed by Ekman (1994). A meta-analysis by Elfenbein and Ambady (2002) identified contempt as the emotion with lowest recognisability based on its facial expression cross-culturally (43.2%). Therefore, if contempt is a basic emotion, then given the definition of basic emotions, its facial expression would be expected to be universal and should have higher recognisability. The evidence for contempt as a basic emotion is not as convincing as for the other six emotions.

Terms for emotional states exist in all languages, as well as a specific concept for each of them. Combined with the evidence presented on distinct patterns for emotions, it makes sense to consider emotions as categories. However, an alternative emotion theory suggests that emotions can be mapped on several dimensions, where overlaps are possible (dimensional approach). For example, according to a dimensional perception, happiness and sadness lie on opposing ends on the dimension of valence (positive vs negative), whereas according to the basic emotion theory happiness and sadness are two distinct systems. Russell (1994) has argued for universality of the dimensions valence and arousal, but questions the universality of emotion categories. According to the dimensional approach, emotions can be mapped on two fundamental dimensions: *valence* and *arousal* (Russell, 1980). An emotion's valence means that emotions can be put on a spectrum ranging from negative (i.e. unpleasant) to positive (i.e. pleasant). Likewise emotions can be judged on a spectrum for arousal, ranging from low (i.e. calm) to high (i.e. excited). It is important to understand which strategy should be employed when people interpret emotions: a categorical or dimensional view. This has been investigated by showing participants emotional stimuli and letting them rate the stimuli on the dimensions valence and arousal. This dimensional answer format reflects a dimensional view of emotions as opposed to providing a list of emotion categories to choose from. The clusters can be represented in the emotional space; two dimensions lead to four squares (see Figure 1). Via multidimensional scaling, a combination of dimensions, similarities and dissimilarities are revealed. For example, sadness is an emotion considered as negative in valence and low in arousal, whereas anger is considered also as negative in valence but high in arousal. This can be seen as confirmed when the participants rate sadness and anger stimuli accordingly leading to separate clusters (Figure 1). Investigations applying multidimensional scaling based on the obtained dimensional ratings have provided evidence for judgements of emotions based on dimensions (e.g. Katsikitis, 1997; Takehara & Suzuki, 1997, 2001).

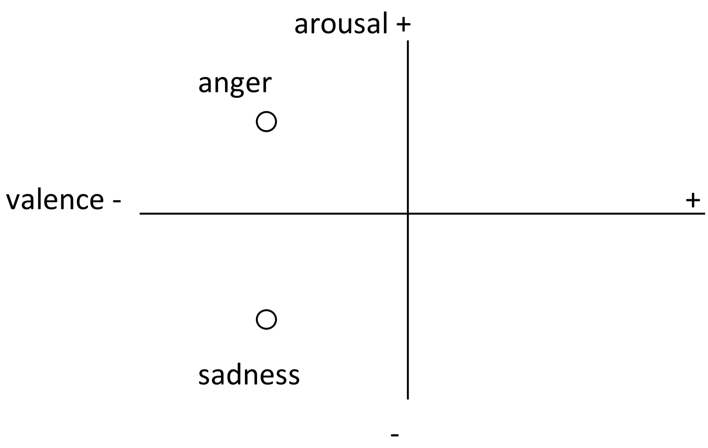


Figure 1. Example for a dimensional classification of emotions.

Research suggests that emotions can be judged categorically and dimensionally. This has led to investigations applying both methods. In such investigations, emotion judgements are made based on categories by assigning stimuli to emotion categories and emotions rated on the dimensions of valence and arousal (e.g. Fujimura, Matsuda, Katahira, Okada, & Okanoya, 2012). Fujimura et al. (2012) systematically investigated the influence of response format (categorical vs. dimensional) on the results. Looking at continuums, the authors tested if category boundaries can be found. To do so, they used emotional blends (e.g. happiness-fear) and created different ratios of those blends for valence (e.g. 25% happiness + 75% fear) and for arousal (e.g. 50% surprise + 50% sadness). The results revealed category boundaries for the emotions shown by a decline in recognition rates at the middle of the valence and arousal continuums based on categorical response format. Furthermore, even when a dimensional response format was applied, Fujimura et al. (2012) found significant category boundaries based on judgements within the dimensional space of valence and arousal for the emotions investigated. It can be assumed, that even when rating stimuli dimensionally, categories are used to recognise emotions. This finding further supports the basic emotion theory of emotions being distinct categories and that the dimensional view is compatible with the categorical view of emotions. Based on the evidence presented here, a categorical understanding of emotions was applied within the current research project.

It should be noted that a distinction between positive and negative emotions is made in the emotion literature on a standard basis. Generally, the distinction positive and negative is made based on the emotion's valence, but is not rooted in an emotion theory. This distinction is also used within this dissertation for the mere purpose of facilitating communication by using categories as well as for brevity instead of listing multiple emotions each time when referring to emotions of positive and negative valence. This differentiation is sometimes made since negative and positive emotions differ for example in their recognisability.

Facial emotion recognition ability.

The recognition of emotions from faces was investigated by early emotion researchers to test the assumption of universality and thereby the existence of basic emotions. Since the existence of the six basic emotions is widely agreed upon, researchers have begun to investigate individual abilities in facial emotion recognition. It is possible to differentiate people's ability to recognise emotions from faces in the laboratory by employing a facial emotion recognition task. A participant is presented with facial emotional expressions and is required to either match the expressions to other pictures of facial emotional expression, to choose a term describing the expressions out of a given list or attribute a term from memory, or to evaluate the expressions on several dimensions either of individual emotions or dimensions like valence and arousal. The categorical vs dimensional answer format reflects a researcher's position on the above outlined emotion theories (categorical vs dimensional view of emotions). The performance of two or more groups at facial emotion recognition can be compared after assessment of facial emotion recognition. A comparison of facial emotion recognition abilities is of interest, since people differ in their ability to recognise emotions from facial expressions (Rozin, Taylor, Ross, Bennett, & Hejmadi, 2005). For example, females seem to be slightly better at facial emotion recognition than males as they achieve higher numbers of correct identifications (literature reviews by JA Hall, 1978, 1990; Kret & De Gelder, 2012). Females have been found to outperform males in facial emotion recognition tasks using static images (e.g. Kirouac & Dore, 1985), brief presentation times bordering subconscious processing (JA Hall & Matsumoto, 2004), and static emotional images of varying expression intensities (JK Hall, Hutton, & Morgan, 2010) as well as emotion blends (e.g. Lee et al., 2013). The literature showing a female advantage at facial emotion recognition varies regarding whether the female advantage is influenced by the type of emotion or applies to all emotional expressions. Specifically, some found a main effect of sex only (e.g. JK Hall et al., 2010) and some found a significant interaction of sex with emotion (e.g. Kirouac & Dore, 1985). That is, the findings on the pattern of the female advantage in facial emotion recognition are inconsistent. It is possible that the inconsistency has its basis in the varying methodology (type of stimuli, task, or answer format) that is generally applied in facial emotion recognition studies.

The most frequently used type of stimulus within facial emotion recognition tasks is static stimuli of high expression intensity. Static stimuli are still photographs/images showing an emotional facial expression. The first standardised picture set of facial emotion is called the 'Pictures of facial affect' (Ekman & Friesen, 1976) and contains black and white still photographs of high intensity facial expressions of the six basic emotions. Since this stimulus set was the first standardised set, it became widely used and is well validated, which is why it is still commonly used today. However, the ecological validity of black and white pictures is rather low, since we generally encounter facial expressions in colour. An example for a coloured set of facial emotion is the NimStim set of facial expressions also including high intensity facial expression of the six basic emotions (Tottenham et al., 2009). There are more stimulus sets of facial emotional expressions (e.g. the Japanese and Caucasian facial expressions of emotion, JACFEE; Matsumoto & Ekman, 1988), but most stimulus sets include facial expressions of high intensity. However, varying degrees of facial emotion expression intensity are encountered in social interactions. Stimulus sets containing such variations in intensity are thus considered more ecologically valid. An example for a stimulus set comprising of static images of facial emotion at varying expression intensities is the 'Facial expression of emotion stimuli and test' (FEEST; Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002), which is based on the ‘Pictures of facial affect’ (Ekman & Friesen, 1976). The varying expression intensities in the FEEST were created by using the morphing technique. Precisely, nine morphed images were created based on the continuum from an image of a neutral facial expression to an image of a high intensity emotional facial expression in steps of 10% (i.e. 10% emotional face, 90% neutral face; 20% emotional face, 80% neutral face etc.). Facial emotional expressions in real-life do not only vary in intensity, they are also dynamic, because they develop and change constantly. Stimuli used for facial emotion recognition testing should reflect this. Static facial images do not necessarily reflect the liveliness and true form of dynamic facial expressions as they occur in daily life (Harwood, Hall, & Shinkfield, 1999). It has been suggested by Collignon et al. (2010) that the diminished ecological validity due to the missing dynamic of expression can influence the outcome of the results. Variations in the type of stimulus applied between facial emotion recognition studies might hence account for variations in the results like in the literature on sex differences in facial emotion recognition.

Inter-individual variability in facial emotion recognition ability becomes especially apparent in clinical populations when compared to healthy individuals. A mental disorder that is characterised by problems in social functioning is ASD. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association (APA), 2013), people with a diagnosis of ASD show restricted, repetitive patterns of behaviour, interests, or activities, and have deficits in social communication and social interaction (including social-emotional reciprocity, non-verbal communication, such as facial expressions, and developing, maintaining, and understanding relationships). Even though a deficit in facial emotion recognition is not specifically stated by the diagnostic criteria, it can be derived thereof. Indeed, literature reviews on assessment of facial emotion recognition in ASD generally conclude a deficit in facial emotion recognition ability in individuals with ASD compared to controls (e.g. Harms, Martin, & Wallace, 2010; Nuske, Vivanti, & Dissanayake, 2013). Baron-Cohen (2002) postulated the extreme male brain theory describing the facial emotion recognition deficit in ASD as an extreme form of the underperformance of typical males compared to females.

Similarly to the literature on sex differences, the results from behavioural facial emotion recognition studies in ASD show inconsistencies. The inconsistencies vary from reports of no impairments in facial emotion recognition in ASD compared to controls (e.g. Piggot et al., 2004) to emotion-specific deficits like in recognition of fear (e.g. meta-analysis by Uljarevic & Hamilton, 2013; Wallace, Coleman, & Bailey, 2008) or recognition of negative emotions (e.g. Ashwin, Chapman, Colle, & Baron-Cohen, 2006). According to Harms et al. (2010), approximately half of the behavioural studies published show impairments in facial emotion recognition in ASD compared to typical developed individuals, whereas neuroscientific and psychophysiological research in ASD consistently shows atypical processing of face emotion stimuli. Harms et al. (2010) suggested that whether or not facial emotion recognition differences can be identified in individuals with high-functioning ASD compared to typical individuals and for which emotions is influenced by the task characteristics (e.g. number of emotions included, intensity of facial expressions, static or dynamic stimuli, matching or labelling task etc.). Considering the number of factors that a researcher has to make decisions on in regards to task characteristics, it is not surprising that the characteristics of the experiments assessing facial emotion recognition vary greatly, contributing to inconsistent findings. Uljarevic and Hamilton (2012) revealed with a meta-analysis that in most behavioural studies investigating visual emotion recognition from face or body in ASD not all of the six basic emotions are investigated. This variation in emotions investigated makes conclusion drawing about emotion-specific deficits difficult. However, Uljarevic and Hamilton (2012) reported a general emotion recognition deficit in the visual modality with an effect size of 0.41 corrected for publication bias and found fear recognition to be marginally impaired in comparison to the recognition of happiness. Inclusion of all six basic emotions in studies on facial emotion recognition in ASD compared to controls is needed to gain insight on the specifics of the facial emotion recognition deficit in ASD. The literature reviews (e.g. Harms et al., 2010) and meta-analysis (Uljarevic & Hamilton, 2012) on facial emotion recognition in ASD show that the variability of applied methodology can be held accountable to some extent for inconsistencies in results on facial emotion recognition ability in ASD. Harms et al. (2010) concluded that some behavioural methodologies are not sensitive enough to capture the differences in facial emotion recognition between individuals with ASD and neurotypical individuals. The sensitivity of a task can be increased by increasing the difficulty in recognising the facial emotional expressions. For example, a short display time and subtler facial expressions of emotion make recognition more difficult, but also contribute to ecological validity. A facial emotion recognition task including these characteristics might be more appropriate to assess facial emotion recognition in high-functioning ASD.

Despite the mixed findings on facial emotion recognition deficits, individuals with ASD self-report problems with reading social cues from faces, e.g. “I do not read subtle emotional cues. I have to learn by trial and error what certain gestures and facial expressions mean.” (Grandin, 1995, p. 156). It seems evident that individuals with ASD have deficits in facial emotion recognition and the social motivation hypothesis has been proposed to explain deficits in emotion processing seen in ASD (Dawson et al., 2002; Dawson, Webb, & McPartland, 2005; Grelotti, Gauthier, & Schultz, 2002). According to this hypothesis, individuals with ASD might find social interactions less rewarding than self-absorbed activities and as a result engage less in social activities throughout their development, which in turn would influence the development of emotion recognition negatively as it is influenced by experience. The ability to recognise others’ facial emotional expressions is a process and develops throughout the whole lifespan, starting in infancy, continuing further through adolescence with peak in adulthood (De Sonneville et al., 2002; Thomas, De Bellis, Graham, & LaBar, 2007). Leppänen and Nelson (2009) suggested, based on a literature review, that social interactions are necessary for the development of facial emotion recognition, as repetitive exposure to emotional facial expressions helps with understanding and decoding. Whereas this hypothesis seems plausible to explain a general emotion recognition deficit often found in ASD, it cannot really explain a deficit in the recognition of specific emotions like the deficit in fear recognition in ASD.

According to the weak central coherence theory, individuals with ASD focus too much on small details and thereby often miss out on the bigger picture (Frith, 1989/2003). Applying this theory to facial emotion recognition in ASD, it might be that individuals with ASD apply featural processing (Happé & Frith, 2006) whereas configural processing is often needed to correctly identify an emotional expression. In configural processing, the relative position of features and their distance to each other are in the focus. This information is used to make judgements about the observed emotion (Ekman & Oster, 1979). In contrast, in featural processing every feature is perceived by its own, in an analytical way. There is evidence that featural processing is sufficient for categorisation of emotions (Cottrell, Dailey, Padgett, & Adolphs, 2001). Presented with clear prototypical facial expressions of basic emotions, a featural analysis might be adequate and therefore applied when decoding these facial expressions (Niedenthal, Mermillod, Maringer, & Hess, 2010) although according to Calder, Young, Keane, and Dean (2000), configural processing is needed at least to some extent. Adolphs (2002) suggested the possibility of applying each strategy depending on the emotion observed. For example, featural processing might be sufficient for recognition of happiness due to the salient smile (Leppänen & Hietanen, 2007), but for recognition of negative emotions, configural processing might have to get applied in addition. Difficulty in decoding facial expressions based on featural processing arises when there is some overlap in the facial movements associated with the production of certain facial emotional expressions. For example, fear and surprise have the overlapping features of the mouth and eyes opened, only distinguishable by the lowered outer eyebrows in fear. The perception of a fearful facial expression as surprised occurs quite frequently; Conson et al. (2013) found a mean confusion rate for fear as surprise of 26%. That means, a specific facial expression can contain cues for different emotions, which makes them ambiguous (Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). This ambiguity makes a precise decoding necessary beyond the analysis of a single feature. Given the featural overlap between some emotional expressions, it is reasonable to assume that featural processing alone is insufficient for such emotions. It is possible that individuals with ASD process faces in a different manner compared to those without ASD in that they rely more on featural than configural processing, which would be in line with the weak central coherence theory (Frith, 1989/2003).

An argument for video stimuli of varying facial emotional expression intensity.

The majority of facial emotion recognition studies is conducted based on static images, which is why emotion researchers lately began to promote the usage of dynamic stimuli. In a recent literature review it was concluded that to advance our understanding further, it is necessary to conduct facial emotion recognition research based on dynamic stimuli (Krumhuber, Kappas, & Manstead, 2013). Using dynamic stimuli for facial emotion recognition assessment can enhance the ecological validity of the obtained results compared to tasks using still images and might lead to a more reliable assessment of facial emotion recognition ability. A method commonly used to create dynamic stimuli is the morphing technique. In many instances of morphing, two pictures are gradually morphed from the one into the other by creating artificial images according to predefined linear increments and the resulting individual images are then played in a sequence following each other to create the dynamic (e.g. Dyck, 2012; Kamachi et al., 2001; Suzuki, Hoshino, & Shigemasu, 2006). It is possible to morph from a neutral facial expression into an emotional expression (and vice versa) or from one emotional expression into another. The advantage of this method is that it is highly standardised, as the number of increments and thereby the number of images (frames) as well as the presentation time of each frame and thereby the exposure time can be kept constant across all sequences. Using the morphing technique, the intensity of the expressed emotion and the emotion itself can be systematically varied. Hence, it is particularly useful when examining sensitivity to emotion perception from faces (e.g. Montagne, Kessels, De Haan, & Perrett, 2007; Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001; Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000). In this case, the participant is presented with a facial expression which develops frame by frame from neutral into an emotional expression (or vice versa) and the participant has to indicate the moment when they perceive the emotional facial expression to appear (onset) or disappear (offset). The frame number at which the participant indicated to have perceived change in emotional content provides information about their perceptual sensitivity and is therefore different from facial emotion recognition.

Unfortunately, the strengths of the morphing technique (highly standardised) lead to some of its limitations. For example, it is questionable that facial features change as linearly as morphing creates these changes and therefore raises concern as to the naturalness of the emotional expression produced (Matsumoto & Hwang, 2014). The onsets of muscular activations differ in timings (see Jack, Garrod, & Schyns, 2014) and so a smile might appear before the eyes begin to wrinkle (Hess, Blairy, & Kleck, 1997) in a display of happiness. Further, a long presentation time combined with a high number of frames leads to a very slow morphing sequence where the facial expression develops in stages (e.g. 500ms x 50 frames = 25sec) and appears highly unnatural. According to Sato and Yoshikawa (2004), morphing sequences longer than two seconds already appear as unnatural to the observer. The perception of naturalistic movement is also influenced by the emotion category. Using the morphing technique, Sato and Yoshikawa (2004) found the speed perceived most natural to differ between emotions, with surprise being considered most natural at a fast pace, and sadness when moving slowly. The pace at which facial emotional expressions are perceived most naturally matches the measured timings by Hara and Kobayashi (as cited in Yoshikawa & Sato, 2008). The authors analysed the time between emotion onset and peak of the six basic emotional expressions from videos and found a range from 330ms for the fastest emotional expression (surprise) to 1,400ms to the slowest emotional expression (sadness) (see also Hoffmann, Traue, Bachmayr, & Kessler, 2010). Pollick, Hill, Calder, and Paterson (2003) investigated the speed of facial expression of multiple encoders and found a range of 466ms between the slowest and fastest encoder. Interestingly, the inter-individual range was larger than the range between the slowest and fastest emotion expressed by the same encoders, which was 208ms. These findings show that there are emotion-specific and inter-individual variations in speed of facial expression of emotion; a natural occurrence that is generally not considered in morphed sequences. In contrast, video recordings capture these variations.

There is evidence that dynamic facial expressions facilitate recognition of the presented emotion due to the inclusion of motion. For example, dynamic stimuli of facial emotion have been used to assess facial emotion recognition compared to static stimuli and it was found that the inclusion of motion (dynamic stimuli) increased recognition rates (Ambadar, Schooler, & Cohn, 2005; Frijda, 1953; Harwood et al., 1999). In addition to motion having a facilitating effect, emotion-specific temporal characteristics seem to be stored in our mental representations of emotions next to their featural characteristics (a mental representation includes knowledge about an object or concept). That emotion-specific temporal characteristics are part of the mental representation of emotions, was demonstrated in a study by Kamachi et al. (2001). They created morphed sequences of varying speed (slow, medium, and fast) and investigated the effects of speed of facial expression on facial emotion recognition. It was found that the recognisability of individual emotions is influenced differently by speed of expression with sadness being better recognised at slow speed and surprise being better recognised at fast speed. That is, a facial expression of surprise is associated with fast facial movements and this expectation is applied during the process of decoding. It can be assumed that the mental representations of facial emotional expressions we use for recognition of such do include both, emotion-specific facial features and their temporal characteristics. However, in morphed sequences the same temporal characteristics are typically applied to all emotional expressions when facial emotion recognition is investigated. If facial expressions are presented at unnatural speeds (slower or faster), this can lead to reduced recognition rates (Bould, Morris, & Wink, 2008; Kamachi et al., 2013; Recio, Schacht, & Sommer, 2013). A way to account for the variability in speed of expression and to increase ecological validity is to utilise video recordings of facial emotional expressions.

The presented advantages of video recordings have sparked the development of video sets of facial emotion where professional actors or untrained participants are filmed whilst displaying prototypical facial emotional expressions (e.g. the Amsterdam Dynamic Facial Expression Set, ADFES; van der Schalk, Hawk, Fischer, & Doosje, 2011; Geneva Multimodal Emotion Portrayals, GEMEP; Bänziger, Mortillaro, & Scherer, 2012; Multimodal Emotion Recognition Test, MERT; Bänziger, Grandjean, & Scherer, 2009; Simon, Craig, Gosselin, Belin, & Rainville, 2008). However, one important feature that is not included in these video stimulus sets and generally not typically included in stimulus sets of facial emotional expressions is variations in expression intensity. Since most facial emotion expression stimulus sets include high intensity facial expressions, the majority of the published research is based on high intensity facial expressions. As a result, our knowledge about the processing of subtler facial emotional expressions is limited.

For the assessment of facial emotion recognition it is important to include varying expression intensities, as in social interactions subtler displays of facial emotion are encountered as well. Ekman and Friesen (1978) suggested intensity ratings when FACS-coding stimuli from trace to maximum, acknowledging the variations in expression intensity seen in facial expressions of emotion. Motley and Camden (1988) compared posed to spontaneous facial expressions of emotion and realised that spontaneous expressions are mostly of low to intermediate intensity. Additionally, it is often claimed that full intensity facial expressions of emotion are the exception in everyday social interactions (Hess, Adams, & Kleck, 2009) and that subtle displays of face emotion are very commonly seen, which makes them a major part of facial emotion recognition (Matsumoto & Hwang, 2014). It is thus surprising that variations in expression intensity are not generally included in video sets of facial emotion. The importance of including variations in facial expression intensity when assessing an individual's ability to recognise emotions from faces becomes even more apparent based on the claim that people generally are not overly competent at recognising subtle expressions (Matsumoto & Hwang, 2011). Such propositions are furthered by research results where static morphed images of varying intensities were applied within a facial emotion recognition experiment and shown that the accuracy rates linearly increased with increasing physical expression intensity (Hess et al., 1997). This finding shows that people are less good at recognising subtle expressions than at recognising high intensity expressions. In addition, being good at recognising emotions from faces displayed at high intensity does not necessarily translate to the recognition of subtler facial emotions. However, this conclusion is based on static images and it is not clear whether subtler displays from video stimuli would also be less recognisable than expression of higher intensity, since dynamics facilitate recognition. Including subtle expressions in emotional stimulus sets would allow for a broader and more reliable assessment of facial emotion recognition in addition to increasing the ecological validity of the results. The knowledge gained from facial emotion recognition ability based on varying expression intensity will advance our knowledge about facial emotional expressions and the recognition thereof. The importance of considering the whole range of emotional expression intensity in video stimuli is thus highlighted.

The discussion of the facial emotion recognition methodologies available has been taken into consideration for the current research project. That is, since static images and morphed sequences are lower in ecological validity, it was decided to use video stimuli for the experiments conducted within this research project on facial emotional expressions. This chapter also showed that it is most common to apply stimuli of prototypical facial expressions at high expression intensity in facial emotion recognition research. We gained great knowledge about facial emotion recognition with this method, although our knowledge about recognition of subtler emotional displays remains rather limited. Furthermore, the ease to correctly label stimuli of high expression intensity with an emotion term makes the detection of any existing group differences in facial emotion recognition ability in typical as well as clinical populations more difficult. The discussion within this chapter suggests that if a facial emotion recognition experiment includes varying intensities of expression and a wider range of emotional expressions, then this should provide a broader and more fine-graded picture about an individual’s facial emotion recognition ability. The demand for research applying video-based face emotion stimuli of varying expression intensity is thus highlighted. However, research can only be conducted applying video-based face emotion stimuli of varying expression intensity if such stimuli are available. To my awareness, no validated video-based face emotion stimulus set based on FACS coding and including varying levels of expression intensity is published. The current research project therefore set out to develop video stimuli of varying expression intensity. Consequently, videos portraying facial emotions at varying levels of intensity of expression were intended to find application in the current research project to investigate facial emotion recognition in the general population and in high-functioning ASD.

Whereas in this chapter facial expressions were presented as an output, i.e. a form of expression providing information about oneself to others, Chapter 2 is a literature review with focus on the production of facial emotional expressions and how these relate to facial emotion recognition as an input function. Precisely, facial mimicry in response to observed facial emotional expressions is presented and its potential function within processing of facial emotional expressions is discussed. Moreover, the influence of explicit imitation of observed facial emotional expressions and facial mimicry on the recognition of facial emotions is discussed.

CHAPTER 2—Facial mimicry, proprioceptive feedback, and facial emotion recognition

In social interactions people tend to mimic each other and mostly without the awareness of doing so; a phenomenon called mimicry (Lakin, Jefferis, Cheng, & Chartrand, 2003). For example, when two people are sitting opposite one another and one person is shaking their foot it is likely that the other person will also shake their foot without the conscious intention to do so (Chartrand & Bargh, 1999). The occurrence of mimicry is well documented, from postures and positions, but also in regards to facial emotional expressions (literature reviews by Chartrand & van Baaren, 2009; Gueguen, Jacob, & Martin, 2009; Hess & Fischer, 2013, 2014). The mimicry in response to observing facial expressions is called *facial mimicry*. Throughout this dissertation where focus is on facial expressions of emotion, the term facial mimicry is used in relation to facial expressions of emotion. The aim of this chapter is to define facial mimicry and to provide a theoretical background on facial mimicry. This chapter further aims to discuss the potential link between face muscle feedback to the brain (that is the result of facial mimicry as well as intentional imitation of observed facial emotional expressions) and the recognition of observed facial emotional expressions underlying the conceptualisation of the current research project.

Facial mimicry.

Hess and Fischer (2014) postulated four characteristics critical for mimicry, which can get applied to facial mimicry in particular:

(1) the expressions of two individuals are congruent, which means that the same facial muscles are activated in an observer that are activated in the encoder,

(2) the expressions are shown in time contingency, which means that the facial mimicry appears immediately after observation,

(3) there is a dependency of the mimicking individual on the mimicked one, which means that the emerging facial mimicry is dependent on what is observed,

and (4) the emotional content expressed is shared between both, which means that facial mimicry is not a reaction to the observed facial emotional expression. That is, congruent muscle activations between observer and encoder as a result of a cognitive response in the observer do not constitute facial mimicry. For example, observing an angry facial expression and reacting with an angry facial expression based on cognitive evaluations of the observed anger (e.g. as being inappropriate) does not constitute facial mimicry. Making a sad face in response to an observed sad face based on a cognitive evaluation of the other's situation and feeling sad for the other person does also not constitute facial mimicry. Likewise, intentional imitation of observed facial expressions does not constitute facial mimicry.

One important distinction has to be made regarding facial mimicry. That is, facial mimicry can be overt (i.e. clearly visible) or covert (not visible to the human eye). Overt mimicry contributes to building rapport between people and liking each other and is dependent on the social context (literature reviews by Chartrand & Lakin, 2013; Hess & Fischer, 2014). For example, smiles are more likely to be mimicked than frowns based on internalised norms (Hinsz & Tomhave, 1991) and probably the anticipation of potential negative consequences displayed frowns can have. In contrast, covert facial mimicry is more basic, a physiological response like change in heart rate in response to emotional stimuli, and occurs also to negative displays of facial emotion. During covert facial mimicry, the muscle contractions underlying are so subtle that no facial movements are visible, which however can be evidenced by electromyogram (EMG) where signals of less than one microvolt from the contracting striate muscle tissue can be recorded via recording disks (electrodes) on the skin (Fridlund & Cacioppo, 1986). For EMG assessment, electrodes are placed on the skin above the muscles that are associated with facial emotional expressions. Guidelines for EMG electrodes placement within face EMG research have been published by Fridlund and Cacioppo (1986), which are in line with the facial actions outlined in the FACS (Ekman & Friesen, 1978; Ekman et al., 2002). These guidelines are still the gold standard today. Since covert facial mimicry is being investigated within the current research project, the term facial mimicry is used from here on referring to covert facial mimicry.

Experimental research applying face EMG has found facial mimicry to happen automatically and spontaneously in healthy people when facing emotional facial expressions without prompting them. For example, Dimberg (1982) investigated facial responses to pictures of facial emotional expressions of anger and happiness that were being observed by using face EMG. He investigated two face muscle sites, the zygomaticus major (involved in smiling) and corrugator supercilii (involved in frowning) and thereby associated with the facial expression of happiness and anger respectively. These two muscles are from here on referred to as zygomaticus and corrugator. The results showed that indeed greater zygomaticus activity was seen in response to happiness expressions and greater corrugator activity was seen in response to anger expressions. Facial mimicry is therefore considered spontaneous and automatic. Facial mimicry also occurs when stimuli are presented so briefly, they are not seen consciously. For example, Dimberg, Thunberg, and Elmehed (2000) presented happy, angry, and neutral faces for 30ms to participants and immediately afterwards neutral facial expressions were presented, which is called backwards masking. This procedure has the effect of the emotional facial expressions to be processed unconsciously. Facial mimicry occurred despite the unconscious processing and is therefore considered an unconscious response. Facial mimicry is also fast occurring. Dimberg and Thunberg (1998) measured facial mimicry via EMG activity in the zygomaticus and corrugator in response to pictures of happy and angry faces across three experiments. They found facial mimicry to occur after 300-400ms of exposure. In live interactions, facial mimicry has been found to appear as fast as within 300ms after stimulus onset (McIntosh, 2006). The characteristics of facial mimicry are thus unconscious, automatic, and fast and that it appears as an affective response to emotional facial expressions.

The occurrence of facial mimicry has been replicated in a multitude of studies measured on zygomaticus and corrugator activity (e.g. Achaibou, Pourtois, Schwartz, & Vuilleumier, 2008; Dimberg & Thunberg, 1998; Lundqvist, 1995; Lundqvist & Dimberg, 1995) or orbicularis oculi (Hess & Blairy, 2001), a muscle also associated with a facial expression of happiness as the activity of this muscle creates the crow's feet around the eyes as seen during Duchenne smiles (Duchenne de Bologne, 1862). However, activity in the zygomaticus and corrugator can also be argued to discriminate on basis of valence, rather than to discriminate between distinct emotions, with zygomaticus activity in response to positive stimuli and corrugator activity in response to negative stimuli. For example, a relationship was found between valence ratings as reported by participants in response to affective stimuli and face EMG activity, with positive valance ratings being associated with activity in the zygomaticus and negative valence ratings with corrugator activity (Larsen, Norris, & Cacioppo, 2003). The deriving implication is that facial mimicry is not a mimicking of face muscle activation of distinct emotions, but a mimicry of the valance of the observed expressions, as has been argued by Hess and Fischer (2013). The evidence for facial mimicry of specific basic emotions based on other muscle sites than zygomaticus and corrugator is limited, as most investigations only include those two muscles or fail to find discernible EMG activity in other muscles in response to associated facial emotional expressions. However, a few studies reported EMG activity in the levator labii (wrinkles the nose) in response to disgust (Lundqvist, 1995; Lundqvist & Dimberg, 1995) and lateralis frontalis activity (pulls the eyebrows up) was reported in response to expressions of fear (Lundqvist, 1995) and surprise (Lundqvist, 1995; Lundqvist & Dimberg, 1995). Reports of facial mimicry of sadness are mostly based on corrugator activity (e.g. Lundqvist & Dimberg, 1995) and not depressor anguli oris (pulls the mouth corners down) even when depressor anguli oris activity is assessed, contrary to what would be expected for the facial expression of sadness. Investigations assessing facial mimicry based on EMG across face muscle sites other than corrugator and zygomaticus are needed to obtain conclusive evidence on whether facial mimicry is valence-based or distinct for the basic emotions. This need is the reason why in the current work face EMG was measured across five muscle sites.

Theory underlying facial mimicry.

Putting aside the debate whether facial mimicry is valence-based or distinct, since the occurrence of facial mimicry in response to emotional facial expressions is reported throughout the literature, the question arises about the function of facial mimicry. A potential role of facial mimicry is that it is part of the decoding process of facial emotion. This assumption is investigated within the current work, but first underlying theories are presented. Mimicry was probably first described by the philosopher Lipps (1907), which he termed 'objective motor mimicry'. According to Lipps (1907) mimicry serves the purpose to make sense of what is observed. For example, observing the facial expression of sadness would lead to a subtle imitation of the sadness expression but without the awareness of doing so. This unawareness and the attentional focus on what is being observed would therefore result in experiencing the emotion as being in the other person; the own feelings are attributed to the other person. Lipps's (1907) view on facial mimicry and recognising emotions from faces is what is now known as the simulation theory.

The simulation theory gained popularity as a response to the theory-theory. According to theory-theory, our ability to interpret others' feelings and intentions is theoretically inferred from available information, e.g. knowledge about the environment, about oneself, and the other's behaviour (Morton, 1980). Whereas theory-theory is an information-based approach, the simulation theory is an approach using the own body. According to Gallese (2003), understanding of others can be achieved via an unconscious simulation of the observed based on the observer's own resources. The observer feels what the encoder is feeling as a result of simulation, similar to what was suggested by Lipps (1907). It cannot be disputed that we often do use available information to interpret our environment as proposed by theory-theory. For recognising an emotional facial expression an observer would analyse the face, consider the features and configurations and match these with prior knowledge. For example, if a smile is detected and smiles are associated with a happy facial expression it can be inferred that the person is happy. That analysing features and configurations can be a successful approach for facial emotion recognition has been discussed in Chapter 1. However, there is compelling evidence from neurosciences for simulation processes that cannot be explained by theory-theory. A large body of research has shown that the same brain areas are activated in an observer than in the person performing an action, as suggested by the simulation theory. First evidence for equivalent brain activity in the observer as in the executer of an action came from brain imaging studies investigating non-human primates. These studies found activation of motor brain areas generally involved in executing hand and mouth movements in macaque monkeys (F5 brain area) observing these actions but without executing them (e.g. di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Fogassi & Ferrari, 2004; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). This is what is called an 'as-if' loop (Damasio, 1994, 1999; Gallese, 2003), because the observed action is not overtly executed, but the motor system activated as if the observed action was executed, next to the expected activity in the visuo-sensory cortex. The resulting question was how it is possible that there is motor activity in the brain although no action is performed. Mirror neurons were suggested to be responsible for such events of paired activations (Blakemore & Frith, 2005; Iacoboni & Dapretto, 2006; Niedenthal, 2007).

Mirror neurons are visuomotor neurons that fire both, when an action is observed and when an action is executed, with the purpose of implicitly making sense of an observed action (Gallese et al., 1996; Gallese, 2003; Rizzolatti et al., 1996). Whereas the existence of mirror neurons seem to be widely accepted for non-human primates, the existence of the mirror neuron system in humans is highly debated (Eigsti, 2013). Most of the support for the existence of the human mirror neuron system comes from studies investigating it using brain imaging or neurophysiology. Evidence for mirror neurons in humans comes from a meta-analysis based on 125 brain imaging studies conducted by Molenberghs, Cunnington, and Mattingley (2012). They found a pattern of activation across studies in areas that consistently have been linked to mirror neuron activity. Compelling evidence for human mirror neurons comes from a study where in-depth electrodes were inserted in human patients with epilepsy conducted by Mukamel, Ekstrom, Kaplan, Iacoboni, and Fried (2010). The participants had to observe and execute a grasping action as well as facial expressions of smiling and frowning. The neural responses of single cells were recorded in both experimental conditions and contrasted. The brain areas investigated in that study where determined by clinical necessity for electrode insertion and therefore limited. Nonetheless, Mukamel et al. (2010) reported that mirror neuron activity was identified in the medial frontal lobe and medial temporal cortex for grasping actions as well as facial expressions based on single cells that were found to fire both, whilst observing and executing an action. From this evidence presented here it can get derived that humans seem to have neurons with mirroring capability.

Mirror neurons are not only evidence in support of the simulation theory, mirror neuron activity could also explain the occurrence of facial mimicry in response to facial emotional expressions based on the simulation in the motor areas of the brain that send signals to the face muscles. If observing facial expressions of emotion activates mirror neurons, resulting in very subtle muscle activity in the observer's face, then these face muscle activations send signals back to the brain (called afferent muscle feedback). These signals from the own body are generally not consciously experienced and are called proprioceptive feedback. It is possible that proprioceptive feedback affects our cognitions. That bodily responses can affect cognitions is the assumption underlying the embodied cognition theories, further called embodiment, which address body-cognition interactions. According to embodied cognition theories, any kind of bodily action, e.g. body posture, gesture, or facial expression, sends signals to the brain via afferent pathways (i.e. from the body to the brain) and influences our cognitions in a bottom-up manner without us realising it. For example, Strack, Martin, and Stepper (1988) showed experimentally that participants holding a pen with their teeth, sticking out of their mouth, and so unconsciously simulating a smile, rated comics as funnier than participants holding a pen with their lips in a way that smiling was prevented. This finding fits with Darwin's (1872/1965) claim that expressing emotions intensifies the emotion just as suppressing softens it. In both experimental conditions of the study by Strack et al. (1988), muscular feedback from the face influenced the attitudes towards the comics. The smile was perceived as elicited by the comics and interpreted as finding the comics funny. This study constitutes one example of how the body's expression of emotion can influence cognitions without being aware of this influence.

The reverse simulation model.

Applying the interaction of body and cognition explicitly to the decoding of facial emotion, the observation of an emotional facial expression would lead to automatic facial mimicry resulting in experiencing the emotion, which then would make the emotion interpretable. As such, mimicry would serve as a means to emotion understanding; a view that has been adopted by Darwin (1872/1965), Izard (1977), and Tomkins (1984), among others. This process to facial emotion recognition was termed 'reverse simulation model' by Goldman and Sripada (2005) constituting a specific model within the simulation theory to explain facial emotion recognition. For example, if somebody is making a sad face, the observer mimics that facial expression and can therefore feel and understand the other's emotional (sad) state. According to the reverse simulation model, facial mimicry makes emotional states of others interpretable due to self-experience. In this scenario, facial mimicry constitutes a necessary component within the facial emotion recognition process.

The reverse simulation theory seems like a plausible explanation for the difficulties individuals with ASD have in regards to facial emotion recognition that have been discussed in Chapter 1. There is reason to believe that these deficits are rooted in a lack or at least diminished production of facial expressions, which is one of the characteristics of ASD (APA, 2013). Since facial expression production is diminished in ASD, it is possible that this extents to the expressiveness on a more subtle level. Precisely, perhaps spontaneous facial mimicry is also decreased in ASD. Indeed, McIntosh, Reichmann-Decker, Winkielman, and Wilbarger (2006) investigated automatic facial mimicry using face EMG in individuals with ASD and controls and found only the control group to show spontaneous facial mimicry responses to facial emotional expressions. The lack of information from facial muscle feedback during decoding of facial expressions of emotion could be underlying the diminished facial emotion recognition in ASD. As facial mimicry is assumed to have its basis in the mirror neuron system, a lack of facial mimicry seems indicative of a malfunctioning mirror neuron system. Based on this assumption, Ramachandran and Oberman (2006) proposed the 'broken mirror hypothesis' to explain the deficits in social-emotional functioning in individuals with ASD. If the function of the mirror neuron system is to facilitate understanding of others, then a lack of mirroring ability would explain diminished understanding of others. Evidence for deficits in mirroring of emotional content in ASD comes from neuroscientific investigations (literature reviews by Perkins, Stokes, McGillivray, & Bittar, 2010; Williams, Whiten, Suddendorf, & Perrett, 2001). For example, a brain imaging study conducted by Dapretto et al. (2006) required children with ASD and controls to intentionally imitate as well as observe facial expressions of emotion whilst brain activity was assessed and the two groups compared. Their findings showed that less brain activation was found in the mirror neuron system of the inferior frontal gyrus in the ASD group compared to controls even though intentional imitation was intact, and that the brain activation was negatively influenced by the magnitude of symptom severity in ASD. This study suggests that individuals with ASD have a malfunctioning mirror neuron system. The malfunctioning mirror neuron system and the resulting lack of facial mimicry could be underlying the deficits in facial emotion recognition in ASD and serve as support for the reverse simulation model.

For the reverse simulation model to be a valid explanation for facial emotion recognition, it is necessary that muscle activations lead to respective feelings. Already Ekman et al. (1983) suggested that the explicit activation of facial muscles involved in emotion expression leads to changes in the autonomic nervous system (heart rate, skin conductance level, skin temperature) via proprioceptive feedback processes and as a result an emotion is experienced. For example, performing a smile would lead to the experience of happiness. Evidence for the influence of the production of facial emotional expression on the actual experience of an emotion comes from a study conducted by Hess, Kappas, McHugo, Lanzetta, and Kleck (1992). Participants were instructed to either feel (to generate the feeling but to keep it inside and not show it) the emotions anger, sadness, happiness, and peacefulness, or to merely express these emotions, or to express and feel the emotions. Self-reports ratings of felt emotions were obtained and showed that even the experimental condition of mere production of facial emotional expression led to emotion experience, despite the instruction to not feel and only express the emotion. This occurrence demonstrates that afferent feedback from muscles can lead to feeling the emotion associated with the expression.

Proprioceptive muscle feedback and facial emotion recognition.

If explicitly expressing an emotion with the body leads to feeling the emotion, then maybe explicit imitation of observed facial expression leads to emotion experience that can be used for decoding of the observed facial expression of emotion. Conson et al. (2013) investigated facial emotion recognition in actors who were trained in two different acting strategies, which they had to apply during the decoding of presented facial emotional expressions. One group acted with aid of the 'Stanislavski Method' where the actor uses contextual information and relates it to their own experiences and feelings in such situations (this relates to theory-theory). The other group used the 'Mimic Method', i.e. the actors explicitly imitate the observed facial emotional expressions and use the resulting generated feeling for decoding (relates to the simulation theory). Both experimental groups and a control group of participants without acting experience completed a facial emotion recognition experiment. The explicit imitation group was found to perform better at recognising the emotions than the Stanislavski group as well as the control group. This study shows that proprioceptive feedback resulting from the explicit imitation of facial emotional expressions facilitates facial emotion recognition. Unfortunately, facial movements were not recorded. Therefore, no information exists about the intensity of the imitation or whether facial mimicry occurred in the Stanislavski and control group. As no data exists on facial mimicry, no statements can be made about the influence of facial mimicry and its proprioceptive feedback on facial emotion recognition; a simultaneous recording of facial EMG thus would have provided useful information.

Since the proprioceptive feedback resulting from explicit imitation of observed facial emotional expressions seems to facilitate the decoding of the observed expressions, it is possible that the intensity of proprioceptive feedback even from facial mimicry (which is much more subtle than intentional imitation) also influences facial emotion recognition. Hess and Blairy (2001) tested this assumption and the reverse simulation model as a whole. They tested experimentally whether facial mimicry is indeed a mediator between experiencing an emotion resulting from observing a facial emotional expression and decoding of observed facial expression of emotion as suggested by the reverse simulation model. Participants were presented with video stimuli of elicited facial emotion (happiness, anger, disgust, and sadness) and had to decode the facial expressions – the encoders in the videos had to imagine or think about emotion-relevant events to elicit the respective emotions from memory. Simultaneous to watching the videos, participants' face EMG was measured from the corrugator, orbicularis oculi, and levator labii. Results showed that facial mimicry was replicated in response to faces portraying happiness, anger, and sadness and the experience of the emotions happiness and sadness was reported by the participants in response to the facial expressions. Hess and Blairy (2001) did not find evidence for an influence of the intensity of facial mimicry on the experience of emotion and neither did they find an influence of the intensity of the experience of emotion on decoding accuracy. They also did not find evidence for a direct link between the intensity of facial mimicry and decoding accuracy. These results seem like compelling evidence against the reverse simulation model (Goldman & Sripada, 2005) or Lipp's (1907) model. However, the important question is as to why the *intensity* of facial mimicry or the *intensity* of the experienced emotion should influence decoding accuracy. None of the simulation models includes intensity as a factor; the focus is on the simulation itself, not the intensity of the simulation. Therefore, to investigate whether facial mimicry is necessary for facial emotion recognition, it would make sense to investigate if the accuracy rates of facial emotion recognition drop when facial mimicry is suppressed.

To my awareness, there is only one study that recorded face EMG and investigated the effects of face movement manipulations and the resulting proprioceptive feedback on facial emotion recognition. Specifically, enhanced proprioceptive feedback was induced that was congruent with the observed facial emotional expression (explicit imitation) as well as facial mimicry suppressed. Schneider, Hempel, and Lynch (2013) presented participants with dynamic morphed sequences of facial emotional expressions developing from a neutral expression to apex of one of the six basic emotions. Participants were split in three groups. One group was instructed to explicitly imitate the facial expressions as they were presented and developed; one group was instructed to suppress any of their own facial movement like when putting on a mask to hide feelings; and one group had no particular instruction regarding facial movements. Face EMG was recorded over three face muscle sites, corrugator, zygomaticus, and levator labii. Participants had to label the presented emotion as soon as they thought to have recognised it, but could change their answer throughout the morphing sequence (duration: 18sec); the final answer was given after the morphed sequence reached its end. (Note that this is a prime example of slow moving morphed sequences as discussed in Chapter 1). The authors found that explicit imitation led to faster accurate recognition of emotion (earlier in the morphed sequence), whereas instructed suppression of facial expressions hampered recognition. Explicit imitation did not lead to an advantage over the no face movement manipulation condition. The authors interpreted the results to show that the proprioceptive feedback accompanying facial mimicry (no face movement manipulation condition) as well as explicit imitation facilitates recognition. However, facial mimicry during the no face movement manipulation was not replicated (but explicit imitation of the observed facial emotional expressions led to discernible face EMG activity). That is, Schneider et al. (2013) did not measure significant increases in EMG activity during stimulus presentation compared to baseline, which however could be a power problem due to the small sample size of 17 participants per group. More likely, the absence of facial mimicry could be the result of using very slow moving morphing sequences, especially as facial mimicry is a fast response (Dimberg & Thunberg, 1998). It should further be noted that suppressing face movement is different to interfering with facial mimicry as different concepts are underlying. Suppressing facial movements is cognitively effortful, as it is based on the intention to hide ones' feelings whereas facial mimicry is an unconscious spontaneous response. Accordingly, suppressing facial expressions does not mean that facial mimicry is also suppressed. Indeed, Schneider et al. (2013) did not find differences in face EMG activity between the suppressing and no face movement manipulation condition for any of the three face muscle sites investigated. Based on this study, it remains unclear as to whether explicit imitation facilitates recognition more than facial mimicry helps decoding of facial expressions and interference with facial mimicry diminishes recognition.

A possibility to examine the effect of absent facial mimicry on facial emotion recognition is to investigate individuals whose facial muscles are paralysed. Neal and Chartrand (2011) conducted a study on a sample of older age females who underwent botulinum toxin Type A (Botox; Allergan, Inc., California) treatment and compared their facial emotion recognition ability to a sample of older age females who received Restylane (Medicis Pharmaceutical Corp., Arizona) injections; both groups got their treatment for cosmetic purposes. Where Botox is injected, the release of the neurotransmitter acetylcholine is blocked. This neurotransmitter needs to be released though to send signals to the brain and vice versa. Hence, the muscle is paralysed (Hennenlotter et al., 2009). Restylane in contrast is affective under the skin, not in the muscle. The acid included in Restylane keeps water in the place where it is injected and serves thereby as filler, leaves the signalling from muscle to brain intact (Brandt & Cazzaniga, 2007; Davis, Senghas, Brandt, & Ochsner, 2010). Since Botox blocks muscle movements, it also blocks facial mimicry, but without adding a cognitive component by instruction or mechanical manipulation. Therefore, this study allowed to investigate the effect of facial mimicry on facial emotion recognition. This study seems to support the assumption that facial mimicry is part of facial emotion recognition, as lowered accuracy rates were found in the Botox group compared to the Restylane group; response latencies were unaffected. These results seem to show that the absence of facial mimicry (due to Botox injections) diminishes recognition. However, it is also possible that not the absence of facial mimicry diminished recognition but that the increased proprioceptive feedback due to Restylane injections increased recognition. The latter interpretation is based on the finding that muscular activity increases when it meets resistance whilst contracting (Vallbo, 1974). Davis et al. (2010) found an enhanced reaction to negative film clips (measured by self-report of valence) after injection of Restylane compared to before. This could be an indicator for the filler (Restylane) to lead to more restriction and therefore increased signalling, which then is also experienced more dominantly. In line with this, Neal and Chartrand (2011) conducted a second experiment where restricting gel was applied on the face in one group and on the arm in the control group and the face-gel group was found to outperform the arm-gel group in facial emotion recognition. This finding supports the assumption of an enhanced muscular feedback to the brain facilitating facial emotion recognition. A control group without any facial treatment would have been necessary in the study by Neal and Chartrand (2011) to learn about the influence of facial mimicry on facial emotion recognition, ideally combined with face EMG.

To investigate the influence of facial mimicry on facial emotion recognition, a study needs to be conducted including in an experimental condition with no face movement manipulation compared to an experimental condition with blocked facial mimicry but without increasing proprioceptive feedback. Accordingly, Rives Bogart and Matsumoto (2010) investigated individuals with face paralysis (Moebius syndrome) and the results of the study suggest that facial mimicry is not necessary for facial emotion recognition. The participants in this study were bilaterally paralysed in their face. That is, no facial mimicry was possible for these participants at all, which also meant that no facial muscle feedback could be send back to the brain. Despite the paralysis, participants performed no different at recognition of facial emotional expressions than the non-paralysed controls. This result would suggest that facial mimicry has no facilitating effect on facial emotion recognition. However, the nerve damage leading to Moebius syndrome supposedly happens prenatal (Briegel, 2006). It is hence possible that the patients learned alternative and compensating strategies in the time facial emotion recognition developed. If alternative and compensating strategies were developed to decode facial expressions of emotions by the patients with face paralysis, then the comparable performance in recognition to controls could be based on the application of those strategies. Since no information is available on whether alternative and compensating strategies were applied or not, this poses a strong limitation to the study and it remains uncertain whether facial mimicry is necessary for facial emotion recognition or not. Investigation in healthy individuals might be more informative as compensatory strategies for emotion recognition from faces should be less likely to be applied.

The reverse simulation model has gained substantial attention within the recent years in regards to facial emotion recognition and the reviewed literature in this chapter showed that there seems to be support for the reverse simulation model and facial mimicry being a component of facial emotion recognition with little counter-evidence. Though, studies are needed investigating the role of facial mimicry within facial emotion recognition by employing measures to assess facial mimicry on a physiological level, because most investigations have not investigated the effects of face movement manipulations on the muscular activity itself. Simultaneous assessment of face EMG would provide invaluable information about the underlying facial feedback within each of the experimental conditions as this provides information about whether face movement manipulations induced proprioceptive feedback that is congruent or incongruent with the displayed facial emotional expression. Based on the methodology that was applied in the studies presented above, it remains unclear whether blocked proprioceptive feedback diminishes recognition or enhanced proprioceptive feedback increases recognition, or both. Nonetheless, it seems evident that proprioceptive feedback does provide information that can be useful when decoding facial emotional expressions. To disentangle the effects, a study needs to be conducted investigating the effects of blocked facial mimicry, enhanced facial feedback, and no face manipulation on facial emotion recognition within one single study to be able to relate the effects to each other, which is one of the aims of the current research project. This chapter also demonstrated that it is necessary to further investigate facial mimicry and whether it is possible to discriminate emotions based on distinct face EMG activity patterns. As already discussed in Chapter 1, most research applied high-intensity stimuli of facial emotional expression and so not much is known about facial mimicry in relation to the intensity of observed expressions of emotion, which is investigated within the current research project. In the next chapter the methodological considerations underlying the conducted studies within this dissertation are discussed, which includes critical discussion about participant choice and recruitment and about the paradigms and measures used.

METHODOLOGY

CHAPTER 3—Critical reflection upon methodology used

Limitations and delimitations surrounding participants and recruitment

The populations of interest for the current research project were the general population and the high-functioning ASD population. Methods commonly applied for recruitment of a general population sample are public advertisement and online advertisement. However, when recruiting from a local area, then participants have to travel to university, which can limit testing sessions to the evenings and weekends for participants with regular jobs. Alternatively, it can be recruited from the university community. When recruiting from the academic environment, online advertisement, advertisement via posters on Campus, and word of mouth are common methods. The success of this method is dependent on people knowing about research participation advertisement and actively seeking them out, i.e. checking noticeboards. As a result, there is a chance that it is the same people participating in the research offered and practice in research participation might influence the results, particularly if research undertaken within a department is similar. It has to be noted that within the current research project, participants were only allowed to participate more than once if prior participation did not provide knowledge about the aims of the study that could alter participants' behaviour and bias the results, or lead to an advantage based on practice. For example, participants could not participate more than once in a facial emotion recognition study, but could participate in one facial emotion recognition study and a study on the judging of the intensity of facial expressions.

When participation is paid for, it is possible that people with specific characteristics volunteer, i.e. people in need for money where the motivation is not driven by the interest in research. This might influence the behaviour of the participant in a way that they might not take participation seriously. At the University of Bath there is the Research Participation Scheme, which allows researchers to recruit first year Psychology Undergraduate students. This scheme is an easy recruitment option and participants do not have to get paid, as first year Psychology students are required to undertake participation hours as part of the Psychology degree and get compensated with course credit, also making them most accessible. This also comes with an obligation to provide possibilities to those students to participate in research. Yet, the motivation to participate might be diminished, as it is a must and not choice. However, the students can choose which studies they wanted to participate in; there always is a choice between multiple studies leading to assume that at least some motivation for compliance within participation should be present. The Research Participation Scheme was used next to word of mouth, posters on campus, and advertisement on the web-noticeboard to recruit participation for the studies of the current research project.

With internationally renowned universities like the University of Bath, the samples generally result in cultural diversity. This might have an influence on the results, as language concepts might differ across cultures and this can have an effect on the data as the facial emotion recognition task was a verbal task; emotion labels had to be assigned. However, for most universities and also for the University of Bath, fluency in English constitutes one of the entry requirements to university. Participant recruitment is a slow process and getting sufficient sample sizes hard to achieve, which is why excluding foreigners from participation cannot be afforded given that research is often limited to a certain time frame. The participation in the studies from the current research project were thus open to all nationalities.

When recruiting from the university community, it is very likely that the majority of the participants are Undergraduate students, as these make for the highest number of people in an academic environment. It is further very likely that this recruitment method leads to a sample with participants predominantly studying from Psychology, since it is part of the Psychology Undergraduate degree to undergo research participation hours. There is the possibility that Psychology students perform differently (probably better) than students from other departments, since they are used to participate in experimental research, which might lead to enhanced performance based on practice. It is further possible that psychology students perform better due to the social component involved in facial emotion recognition. Seeking out psychology as the area of studies might be associated with an increased interest in people and social interactions, which might lead to higher motivation to perform well, or to better performance based on personality traits, e.g. empathy. This can lead to variance within the sample and influence the results. Generally, with participants from an academic environment, the results might not generalise to the general population as a whole, as factors like IQ might also influence the results. To retrieve a little more of a general population sample, recruitment can be widened to students and staff (including academic and administrative staff). Recruitment from the academic environment is general practise within psychological research, as these people are accessible, do not have to travel for participation and are therefore not compensated for their travel costs, which is an important factor when financial resources are restricted. Further, these people are interested in the research undertaken at their university leading to motivation to be a supportive part of that. The non-clinical participants for the studies conducted within the current research project were recruited from the University of Bath community, including both, students and staff.

Access to clinical populations is generally limited for researchers not associated with a clinical institution, which makes recruitment of clinical populations difficult. One recruitment option are events that are held specific for a certain population. The University of Bath runs a yearly 3-day Autism Summer School for individuals with high-functioning ASD who are either already studying at university or about to start university. Such events offer the possibility for data collection, although data can only be collected for the duration of the event and to times that do not interfere with the event itself. The ASD samples for the current research project were recruited from the Autism Summer School.

The research conducted within this dissertation focusses on individuals with ASD that are high-functioning. Given ASD constitutes a spectrum ranging into low functionality, generalizability of the results obtained to the lower functioning end of ASD is not assured. Most published studies are based on high-functioning ASD (Winkielman, McIntosh, & Oberman, 2009) and there are several ethical and practical reasons for this. Research participation requires participants to focus and concentrate on a task for a fixed certain amount of time, for which a certain level of cognitive abilities are required, both of which is difficult with low-functioning individuals with ASD. For these reasons, only high-functioning individuals with ASD were recruited for the ASD studies within this dissertation.

Limitations and delimitations of lab-based assessment of facial emotion recognition

In case respective facilitates are available, then data assessment can be carried out for several participants simultaneously. This procedure saves time, but might influence some participants and their ability to focus and concentrate. If more than one participant is tested at the same time, distraction can be minimised by starting the testing sessions all at the same time and have participants wear headphones. When the experimenter is in the room with the participant, this can have an effect on participants' behaviour by either distracting them or altering their performance and responses. However, some laboratory facilities do not offer the option to leave the room whilst staying in contact with the participant in case they have questions, there are technical issues, or participants want to withdraw from the study. For the assessment behavioural data within the current research project sometimes the data of several participants was collected simultaneously with between one and seven participants and the experimenter was always in the room.

When testing sessions are conducted in research laboratories then that constitutes an artificial environment and participants are asked to engage in artificial and specific tasks, which might alter participants' behaviour, decreases generalisability, and challenges ecological validity. However, experiments allow for replication and the quantitative data obtained is analysable in inferential statistics. Also when assessing facial emotion recognition in a laboratory, this does not resemble a social interaction as people generally encounter it. Most facial emotion recognition research is conducted as a computer-based experiment meaning that the interaction is unidirectional for participants unlike in social interaction outside the laboratory. It has been shown that modifications of an experimenter's eye contact influenced participants' eye movements only when interacting live, not when video recordings were used (Freeth, Foulsham, & Kingstone, 2013). This shows that behaviour is different when conducting a computer task compared to interacting with people. The participants' awareness that facial emotion recognition is assessed might influence their behaviour in that they might show different face scan pattern than in social interactions outside the laboratory or focus more on the emotional cues than they normally would. For example, in live interactions the eye gaze changes from being addressed to the face to away from the face, which affects recognition by missing certain emotional cues.

Within a lab setting, participants are required to stare at a face. In addition, motivation to perform well is often increased and the attention more focussed. It has even be claimed that group differences in facial emotion recognition experiments are driven by motivation (Buck, 1988). However, experiments carried out in laboratories allow for greater control, as variables can be manipulated whilst others are kept constant, which also allows for the establishment of cause and effect. That is, if all relevant variables have been controlled, which can be difficult where there are a multitude of influencing factors and the experimenter might not always be aware of all of them. The possibility of controlling factors make experiments a useful method to assess facial emotion recognition, as the results can be attributed to the information that was made available to the participant, i.e. emotional information from faces. Facial emotion recognition was hence assessed within a laboratory setting for the studies presented within this dissertation.

As has been discussed in Chapter 1, there are several types of face emotion stimuli that find application in computer-based facial emotion recognition tasks, ranging from static images to dynamic stimuli, the latter either in form of morphed sequences or video recordings. Of these, videos inherit the highest ecological validity as long as they have not been manipulated by deleting frames or altering the speed with which the frames are presented. However, even videos do not resemble live social interactions. In face-to-face social interactions emotional information is available from several modalities, e.g. auditory, context, body movements etc. However, since the ability to recognise emotions from faces is of interest within this dissertation, it is necessary to exclude all information from other modalities than the face. Face emotion stimuli are different to facial expressions that are encountered in social interactions. The stimuli usually show a full frontal view, whereas in social interactions facial expressions are seen from all kinds of angles. Facial expressions in everyday life are embedded in a wider context, appear spontaneously and are fleeting, whereas face emotion stimuli are usually posed (either by trained actors or lay-participants) and show prototypical expressions. Using prototypes is important in terms of standardisation, but it needs to be kept in mind that spontaneous expressions might not look like prototypes. Prototypical displays of facial emotion were used within the current research project, but from video stimuli.

There are several methods to assess facial emotion recognition, but all tasks have in common that the participant is presented with facial emotional expressions displayed as static images or dynamic sequences and the participant has to judge the emotion in some way. The answer format is determined by the task. For example, a matching task requires the participant to match the stimulus to one of the given pictures of facial emotional expression. Therefore, this particular task requires the usage of static stimuli. A matching task might not actually assess facial emotion recognition, as a mere differentiation between similar and dissimilar emotional images is sufficient to conquer a matching task. In contrast, recognition of an emotion requires additional knowledge that is not contained in visible features (Adolphs, 2002). That is, recognition entails knowing which emotion is expressed in the face and not only whether two facial expressions differ from each other. A labelling task requires the participant to attribute an emotion term (label) to the facial expression seen. Since labelling tasks require more recognition than discrimination, a labelling task was applied throughout the studies presented in this dissertation.

There are two forms of a labelling task answer format, either the participant chooses a label out of provided labels (multiple-choice) or labels freely from memory. There are certain consequences from applying a forced-choice/multiple-choice answer format. Since the given answer choices are based on pre-selection, it is possible that the term the participant had in mind is not contained in the list and has to choose the 'best' option provided. Further, it is possible that people might choose the right label by chance occurrence. Providing only a couple answer choices might not adequately assess facial emotion recognition, as the correct answer might be identifiable via a process of elimination of the most unlikely answers. For example, if happiness is the only positively valenced emotion displayed, it is possible to identify the emotion accurately by recognising its valence, i.e. categorising into negative and positive. As a consequence, the consensus among participants may increase artificially (Russell, 1993, 1994). It is therefore beneficial to provide a multitude of answer choices and include more than one positive emotions (e.g. pride). Therefore, complex emotions were applied within the current research project in addition to the six basic emotions to increase the number of positive emotions included.

Some researchers apply an open answer format. The effects of answering formats on accuracy rates within emotion recognition was investigated by some researchers. A freely labelling emotion recognition task was applied and it was found that even though participants did not assign the exact same labels to the stimuli, they chose labels of the same emotion family of the stimuli (e.g. irritation, frustration, etc., belonging to the anger family) and generally the emotion words used were the same as the ones given in forced-choice tasks (Boucher & Carlson, 1980; Haidt & Keltner, 1999; Izard, 1971; Rosenberg & Ekman, 1994). A recent study (Limbrecht-Ecklundt et al., 2013) provides further evidence for forced-choice formats leading to similar results as open-response formats for the basic emotions. The authors consider the use of forced-choice formats in facial emotion recognition experiments as unproblematic. Widen, Christy, Hewett, and Russell (2011) investigated four complex emotions and the six basic emotions and found the accuracy rates from the forced-choice format to be comparable to the freely labelling approach for the six basic emotions. The accuracy rates for the complex emotions were moderate using the forced-choice format, but were low in the freely labelling paradigm. It seems like the emotion categories included in the task influence the applicability of the answer format. The variance in responses between participants from freely labelling answer formats can make analysis difficult. A forced choice format was hence applied throughout the studies presented in this dissertation.

The main dependent variable of interest within current research project was accuracy of response. Accuracy refers to the percentage correct out of the total number of trials for each category. When conducting a facial emotion recognition experiment, the researcher determines which responses by the participants are considered correct and which false. This raises the question as to what constitutes accuracy. There actually is no correct answer, rather it can be said that there is a consensus among observers. Another dependent variable was response time in recognising emotions, which refers to the time participants took to respond from the moment the answer screen was presented until the participant clicked the mouse on their answer choice. Response time is not identical to reaction time. The response times obtained are not generalizable to face-to-face social interactions, as it is possible that participants want to impress with speed or shorten the duration of the experiment by fast responding, or take their time with responding to avoid errors based on speed. Either way, these motivations do not apply in everyday social interactions. To counter this, participants in the studies within the current research project were instructed to imagine being in a social interaction with the people whose faces they observe and to respond in an appropriate speed. This aimed to avoid false responses due to fast responding (speed-accuracy trade-off) and application of additional decoding strategies.

Limitations and delimitations of physiological measures

Emotion processing cannot only be assessed on a behavioural level, but also on a more objective level by taking physiological measures. One such measure is the electromyogram (EMG), which can be used to assess participants' facial movements in response to stimuli. The advantage of taking physiological measures is that participants cannot easily influence or alter these measures and participants can be kept in the dark about the true purpose of the assessment. However, electrodes need to get attached to the participant. It is possible that the electrodes make participants uncomfortable or that they feel restricted, both of which might still alter their responses. Though, high acceptance has been reported towards surface electrodes in the face, on chest, and on fingers (Wingenbach, 2010). It is also possible that participants are a little scared if they have never encountered electrodes before, although surface electrodes are non-invasive and no harm is involved. Nonetheless, electrodes allow to take measures that would not be possible with the eye only, even though it is a time consuming procedure. When assessing face EMG with surface electrodes, it has to be accepted that the measurements are not very specific. That is, activity from a region is measured instead of a certain spot. Needle electrodes are more location-specific, however, since they get inserted in the muscle they are highly invasive. When using surface EMG electrodes it is more difficult to map the data to a specific emotion, even though specific muscle activations have been associated with specific facial features (Ekman & Friesen, 1975) and there are placement guidelines for face EMG (e.g. Fridlund & Cacioppo, 1986). Another issue that can arise from using EMG surface electrodes is that depending on the participant's skin and facial hair, contact between the electrode and the skin might be diminished, and depending on fatty tissue the signals obtained might be less ideal. It is therefore beneficial to ask male participants to shave their facial hair before participation.

Limitations and delimitations surrounding self-report measures

Measures that rely on participants' self-reports are subjective in nature and with that prone to bias. Problems can arise in the case that a participant answers in a socially desirable way. However, even if a participant answers honestly, there is the possibility that they lack the required introspective ability to answer accurately. In addition, the interpretation of the same item can differ between people. An example constitutes the item from the AQ (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) asking individuals about their interest in dates, which can be interpreted as calendar dates or romantic dates. Furthermore, with a Likert-scale type of responding the interpretation of the extent might also differ between people, e.g. a 5 for one is the same as a 3 for another. Then there are response tendencies with some people tending to respond around the middle of a response scale and some people at the extremes. With such response tendencies differences between participants emerge that do not reflect what the measure was intended to assess. Nonetheless, some of these limitations can get addressed. For example, response bias can be addressed by reversing some of the items which inverses the response on the scale. Dishonest responses can get addressed by assuring confidentiality and anonymity. The experimenter should also not watch the participant whilst responding. Misinterpretation of items can get addressed by the experimenter making themselves available for the participant to clarify any item meanings. It should be noted that self-report measures undergo rigorous checks for reliability and validity during the development and after publication often leading to the production of norm data. The fulfilment of these criteria assures that the measures produce consistent results and measure what they are supposed to measure. Furthermore, self-report measure do offer advantages as well; administering self-report measures is a quick and easy way of assessment that also comes with low costs. A self-report measure that was applied within the current research project is the Self-Assessment Manikins (SAM; Bradley & Lang, 1994) to assess participants' affective state. SAM is a picture-based non-verbal measure on a 5-point Likert scale assessing *valence* (1 = very positive, 2 = positive, 3 = neutral, 4 = negative, 5 = very negative) and *arousal* (1 = very weak, 2 = weak, 3 = moderately, 4 = strong, 5 = very strong) on one item each; see Figure 2. One of the advantages of the SAM is that it can be completed within a few seconds. Furthermore, the use of pictures does not require verbalisation that is dependent on semantic concepts. However, the visual portrayal is not intuitive for everyone and therefore verbalisations of the scales were used in addition to the pictures within this research project (i.e. “How are you feeling right now?”).

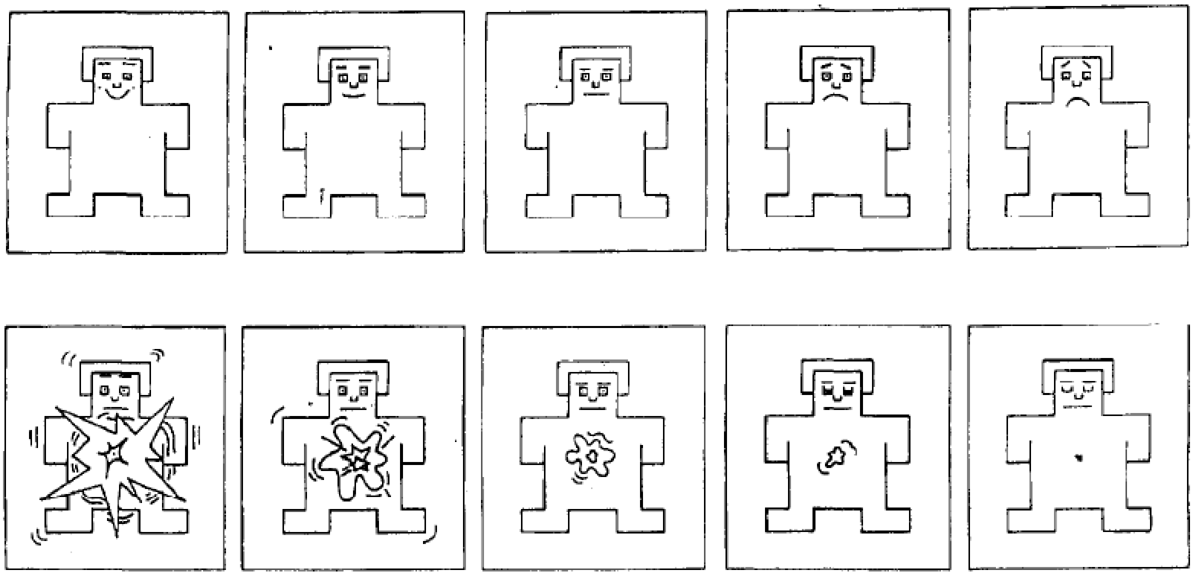


Figure 2. SAM for the assessment of valence (upper row) and arousal (lower row) (Bradley & Lang, 1994).

Having discussed the methodological limitations and delimitations of the methodology applied in the current research project with this chapter, the following chapters present the experimental research undertaken within this research project. The development of video stimuli portraying facial emotional expressions at varying degrees of expression intensity is presented in the following chapter. To assure that the facial emotions from the developed stimuli would not be too difficult to recognise, a facial emotion recognition experiment was conducted on a sample of individuals with high-functioning ASD, which is also presented in the following chapter.

EXPERIMENTAL STUDIES

CHAPTER 4—Development of facial emotional expression videos including varying levels of intensity

Introduction

Following on from the discussion in Chapter 1 that stimuli of higher ecological validity for application within a computer task of facial emotion recognition would be dynamic and would also show facial emotional expressions at varying levels of intensity, it was decided to utilise such face emotion stimuli within the current research project. With most investigations applying static stimuli or stimuli portraying facial emotion at high intensity, there is not only a need to utilise dynamic stimuli at varying levels of intensity, but there is also a lack of such stimulus sets. Whereas a multitude of face emotion stimulus sets including static images exist, a literature search revealed that there are (to my awareness) four stimulus sets published including dynamic facial emotion stimuli of varying intensity levels of expression. A computer-generated stimulus set for the six basic emotions at varying intensities has been published called the Emotion Recognition Task (ERT; Montagne et al., 2007) and three video stimulus sets of that kind; the Database of Facial Expressions (DaFEx; Battocchi, Pianesi, & Goren-Bar, 2005), the MPI Facial Expression Database (Kaulard, Cunningham, Bulthoff, & Wallraven, 2012), and an unnamed set developed by Simon et al. (2008).

Since it was decided to utilise video stimuli of facial expressions for the current research project, the computer-generated ERT was not a suitable stimulus set for this research. The DaFEx (Battocchi et al., 2005) includes three intensity levels of expression of the six basic emotions, however, is limited in the use for emotion recognition research from faces, as emotions are expressed not only facially but also with body posture providing further cues useful for decoding. Additionally, the stimuli vary in length by up to 20 seconds. The MPI contains 55 facial expressions of cognitive states (e.g. remembering) and five basic emotions with context variations at two intensity levels each. However, in the MPI the encoders wear a black hat with several green lights on for head tracking purposes which also covers facial features generally visible in social interactions (e.g. hair) – only the face is visible – all of which lowers ecological validity. Unfortunately, validation data have only been published for the high intensity expressions, not the low intensity. It also has to be mentioned that none of the two video sets of facial emotion expressions have been standardised by FACS-coding. The video set by Simon et al. (2008) consists of videos of the six basic emotions, neutral, and pain posed by eight encoders (4 female, 4 male). The videos are short in length (1 sec), in line with the natural occurrence of facial emotional expressions (Pollick et al., 2003). The advantage of this set is that it is FACS coded and different intensity levels were produced. However, the low intensity expressions are neither published nor validated and access would have been possible only after December 2013 (Simon D. Personal correspondence. 27 March 2013); too late for this study. To date, there is no published fully validated stimulus set containing video recordings of facial emotional expressions including varying intensities of expressions. This highlights the importance to create such a stimulus set.

It was decided to create video stimuli of varying intensities based on an existing video set by extracting short sequences from each video (explained in detail in the Method section below). As a consequence of this method, there was the possibility that the videos' recognisability would be hampered. This possibility is based on the fact that facial muscle activations are independent from each other and therefore can vary in timings (Jack et al., 2014). That is, even a group of AUs that make for a specific emotion can have differing onsets. For example, when expressing happiness with the face then at a very low intensity of expression the wrinkling around the eyes is not visible, but the smile is (Hess et al., 1997), which would mean that the mouth corners are pulled up before the wrinkling around the eyes appears. This also means that by extracting sequences from videos the amount of available emotion cues for decoding of the facial expression is diminished compared to the full videos. The arising question therefore is whether or not the facial emotions would still recognisable from extracted sequences. This question is especially of importance for sequences displaying low intensity expression, since here the most emotional information would be 'lost'.

A measure reflecting the recognisability of an item (here: videos) is the item difficulty index (Crocker & Algina, 1986). The item difficulty index is abbreviated as *p* value, which is not to be confused with the significance level *p*-value. The *p* for the item difficulty index stands for the proportion of correct responses. Precisely, it reflects the percentage of participants giving the correct answer to an item and is calculated by dividing the number of participants that answered correctly by the number of participants in total. The retrieved *p* values can vary between .00 and 1.00. The larger the *p* value, the more participants answered correctly to the respective item (Adkins Wood, 1960). Item difficulty influences the variability of scores and as a result the ability to discriminate participants based on their performance (Thorndike, Cunningham, Thorndike, & Hagen, 1991). Item difficulties of either .00 or 1.00 can be considered meaningless, as they do not have discriminative ability. Further, in multiple choice tasks there is always the possibility of responding correctly by choosing the right answer by chance, and any item yielding *p* values that reflects the chance level of responding of a task (or below) can therefore be considered equally meaningless. If all items of a task yield very high *p* values (e.g. > .95), then the task will, as a result, not have discriminative ability. Therefore, a task should include items with a range of difficulties. The recognition for stimuli of facial emotion should be rather high. That is, the agreement among judges has to be high to ascertain that the intended emotion is portrayed. Therefore, a stimulus set of facial emotion should include items with a difficulty index for each video greater than chance level of responding and smaller than 100%, but with most *p* values occurring at 80-95%. As a result, the stimulus set includes mainly videos of high recognisability, but also includes difficult items to add discriminative ability.

For the current research project it was of importance that the developed video stimuli would have discriminative ability, since it was part of this research project to discriminate participants on their facial emotion recognition ability. For example, the facial emotion recognition ability of males was compared to females (Chapter 6), and individuals with high-functioning autism were compared to controls (Chapter 8). The purpose of the pilot studies presented in this chapter was to develop video-based facial emotion recognition stimuli including varying levels of intensity that would be applied throughout the studies within this dissertation. A first evaluation of the developed videos on basis of the item difficulty index was hence necessary (Pilot Study 1). Following this first evaluation, the stimuli were applied to a sample of individuals with high-functioning autism to test their applicability to a sample with difficulties in nonverbal communication (Pilot Study 2). This was of importance to make sure that the task was not too difficult for this (or similar) populations, to allow application of the stimuli throughout the research conducted within the current research project. Even if the newly created videos were well recognised by a postgraduate student sample, it would be uncertain whether or not this would be the case in a sample of individuals who struggle with emotion recognition from faces (i.e. ASD; see Chapter 1). It was imperative for the videos to be recognisable for individuals with a diagnosis of autism, as facial emotion recognition in the ASD was a component of this research project. Therefore, another pilot study was carried out to test for applicability on a sample of individuals with high-functioning ASD.

Aims and hypotheses

The aims of the present study were to:

(1) select a video stimulus set.
(2) create video stimuli of varying intensities of expression on the basis of the stimulus set selected.
(3) evaluate those newly created videos regarding recognisability on single item level.
(4) test if individuals with high-functioning autism would be able to undertake the experiment measured on their accuracy of response rates for the emotion categories included in the facial emotion recognition task.

It was hypothesised:

(1) that the *p* value of each of the videos would be greater than .10 and smaller than 1.00 and with the majority of videos having a *p* value of .80 - .95, since videos that are very hard to recognise (with a *p* value at and below chance) and videos that are very easy to recognise (with a *p* value of 1.00) are of no value.
(2) for the experiment to be suitable for application in an ASD sample, the accuracy rates for each emotion category as well as for each emotion at each intensity level were expected to be significantly higher than chance level of responding.

Methods

Participants*.*

Recruitment for Pilot Study 1 took place through advertisement on campus in form of posters, word of mouth, and through the web-noticeboard. The sample consisted of 20 participants (10 male, 10 female), aged between 21 and 39 (M = 28.6, SD = 4.7). Postgraduate students of the University of Bath from all departments were invited to take part. Participants were from the Departments: Psychology (n = 6), Political Sciences (n = 4), Mechanical Engineering (n = 3), Health (n = 3), Physics (n = 2), Management (n = 2). The study resulted in an ethnically diverse sample: British (n = 6), Greek (n = 3), Mexican (n = 2), Iranian (n = 2), Indian (n = 2), Serbian (n = 1), Thai (n = 1), Cypriot (n = 1), South Korean (n = 1), American (n = 1). All participants were assumed to be fluent in English, as this is one of the entry requirements to study at the University of Bath. All participants had normal vision or corrected-to-normal vision to assure that the visual emotion cues can be perceived. None of the participants reported a clinical diagnosis of a mental disorder.

Participants for the Pilot Study 2 were recruited from the Autism Summer School at the University of Bath. The total sample consisted of 10 participants (8 male, 2 female) of which 6 (all male) completed the study. Access to clinical populations is generally limited for researchers not associated with a clinical institution, but the University of Bath runs a yearly 3-day Autism Summer School for students from age 16 to 24 with high-functioning ASD who are either already studying at university or about to start university. The participants for the current study were hence recruited from the summer school. All participants had a formal and current diagnosis of an autism spectrum disorder (high-functioning) given by a clinical professional at some stage in their lives, which they had to bring to the Autism Summer School. Participants had normal or corrected-to-normal vision, which is important so that a visual impairment does not hinder the extraction of the emotion cues from the faces. Mean age was 18.2 years (SD = 1.7), ranging from 16 to 21 years. Ethical approval for both pilot studies was given by the University of Bath Psychology Ethics Committee (Reference number 13-044). Participation in these pilot studies was not monetarily compensated.

Stimuli selection.

A literature and general internet search was conducted to find a suitable face emotion video stimulus set of either posed or elicited expressions to edit to varying expression intensities for application in the research of the current research project. Since it was aimed to apply stimuli of higher ecological validity, the following criteria were applied:

- video recordings
- emotions expressed by real people
- encoders looking face-forward into the camera
- videos in colour
- showing only one person at a time
- including a neutral sequence, i.e. blank stare
- showing facial expressions developing from neutral to an emotion
- including the six basic emotions
- including complex emotions and/or different expression intensities
- large number of encoders per emotion included

These criteria led to a preselection of video stimulus sets for visual inspection. Access to stimulus sets was obtained from the developers of the 'Geneva Multimodal Emotion Portrayals' (GEMEP Corpus; Bänziger et al., 2012), the 'Database of Human Facial Expressions' (DaFEx; Battocchi et al., 2005), the 'Derek Harper Database, Texas' (O'Toole et al., 2005), and the 'Amsterdam Dynamic Facial Expression Set' (ADFES; van der Schalk et al., 2011). Since within the current research project emotion recognition from faces was investigated, it was necessary to apply face stimuli that are not confounded with other modalities and have been produced in a standardised manner. The video stimulus sets, which were gained access to, were hence inspected based on the following criteria to select a suitable stimulus set for the current research project:

- having the whole head, but not the rest of the body of the encoder visible in the videos
- expressing the emotions facially only, no body movements
- have no utterances included to avoid distraction and unwanted further cues to the emotion
- having each encoder express all emotions included

- standardisation of the encoders' emotional expressions based on FACS by certified coders
- similar duration of the videos

After those criteria have been applied, the ADFES (van der Schalk et al., 2011) was identified as a suitable stimulus set. The ADFES includes 10 expressions (the 6 basic emotions and contempt, pride, embarrassment, as well as neutral) posed by 22 encoders (12 Northern Europeans, 10 Mediterranean). The inclusion of complex emotions is an advantage of this set. The ADFES is FACS-coded and validated by its developers. This assures that the posed facial emotional expressions align with the prototypical displays. However, all emotions were expressed at full intensity. With a duration of approximately 5.5 seconds per video, it was possible to extract several short sequences to create varying expression intensity levels. Permission to adapt the videos was obtained from one of the developers of the set (Fischer A. Personal communication. 19 April 2013).

Stimuli development.

The 12 Northern European forward-facing encoders of the ADFES were edited. The aim was to create three videos for each of the 120 original videos, one showing a low intensity of the displayed facial expression, one an intermediate, and one a high intensity expression. The different expression intensities were to reflect the spatial distance of the facial features in comparison to their location within the neutral expression, based on the degree of contraction of the relevant facial muscles for the emotional category. The category of 'low' intensity included subtle expressions with a slight activation of facial action units underlying the respective emotional facial expression, i.e. the onset of the facial emotional expression. The 'high' intensity category included the apex of the emotional expressions within the videos, which involved the maximal contraction of all the relevant facial action units and the greatest spatial distance of facial features from where they appeared in the neutral expression. The 'intermediate' category was chosen to be as distinct as possible from the low and high intensity points in the videos, with the muscle contraction and movement of the facial features being mid-point between those intensities (see Figure 3 for an example).

Figure 3. Example of the three created intensity levels of expression. Neutral expression, last frame of fear at low, intermediate, and high expression intensity (from left to right).

To create the videos, a double-editing process was applied. In a first step, the duration of the original videos of 5.5 seconds was shortened in “Windows Live Movie Maker”. That is, certain time periods were selected, but had to start at 0ms to not distort the video based on software specifications. For each original video, a number of sequences were extracted, e.g. 0-1,200ms, 0-1,500ms, and 0-1,800ms and saved as .wmv files. In a second step, each video was individually exported into and edited using “Corel Video Studio Essentials X4”. This software allowed selection by frame and thereby frame-by-frame editing to set the first and last frames of each new video.

For each new video, the desired end frame for each stage of expression corresponding to the appropriate intensity level (e.g. low, intermediate, and high) was selected as the ending point of the video, and then the 25 prior consecutive frames were selected to create the rest of the video. This created videos of equal length and time (26 frames) but with the last frame corresponding to the relevant level of emotional intensity. The original ADFES video format is .mpeg with a frame width of 720 and a height of 576, and 25 frames per second. Every video was saved as .wmv file in High Definition quality with the same frame rate of 25 frames/sec resulting in a duration of 1040ms. Since apex of facial expression is generally reached within 1 second for basic emotions (Hara & Kobayashi, 1995 as cited by Yoshikawa & Sato, 2008), this timing allowed for all videos to start with a neutral facial expression and to continue until the end of the expression. This is more ecologically valid, since outside the laboratory people generally get to see neutral expressions as a point of reference (Matsumoto & Hwang, 2014). It was possible to have the videos start with a neutral facial expression, i.e. blank stare, as the original ADFES videos have a 500ms blank stare before initiating an emotional display of the facial expressions. In the end, videos were created for each of the 12 encoders expressing the 9 emotions at the three intensity levels. Three different sequences were also extracted from the neutral facial expression videos to obtain an equal number of videos per category. A total of 360 videos were created from the Northern European subset of the ADFES. In addition, one video for each emotion at high intensity was created from a female actor from the Mediterranean sample of the ADFES to be used for the practise trials before the experiment.

It has to be mentioned that the duration for which emotional content is visible in the created videos, was determined by the intensity of the facial expression. This means for this new stimulus set that the low intensity videos show a subtle facial expression (i.e. a little twitch), only visible for a very short time (e.g. 200ms). The intermediate intensity videos show emotional expressions a little bit more developed and hence for a slightly longer time. Respectively, the high intensity videos show an almost fully to fully developed emotional facial expression and as the whole development is visible, the emotional period is longer than the blank stare period (see Figure 4). This is due to the naturalistic manner of this stimuli set, where neither the frames nor the frame rate were manipulated.

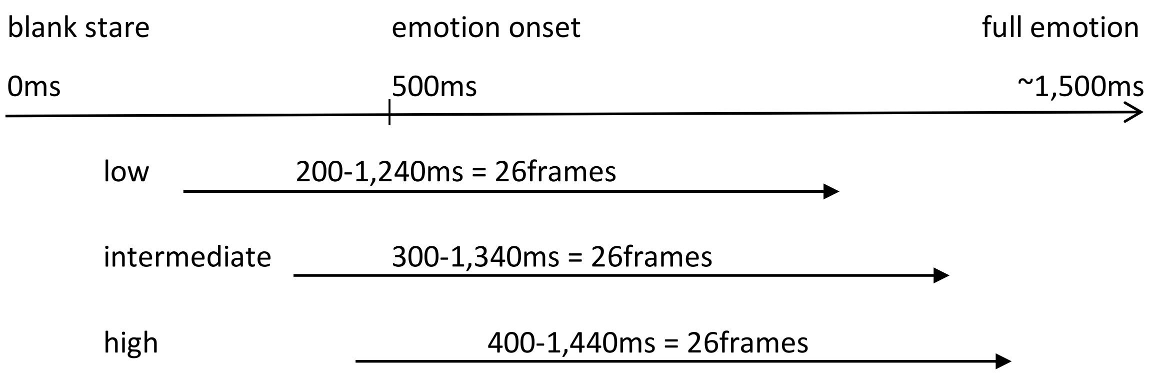


Figure 4. Example of extracted sequences for the created expression intensity levels. The upper arrow shows the first 1.5 seconds of the original ADFES. The three shorter arrows represent examples for a video at low, intermediate, and high expression intensity.

Facial emotion recognition experiment.

The psychological stimuli presentation software E-Prime 2.0 (Psychology Software Tools, 2012) was used for creation (and presentation) of the facial emotion recognition experiment. A labelling task with forced-choice answer format was created, with multiple answer choices given, i.e. the 10 facial expressions included in the ADFES (anger, contempt, disgust, embarrassment, fear, happiness, neutral, pride, sadness, surprise). The answer screen contained 10 response fields including one emotion label each (anger, contempt, disgust, embarrassment, fear, joy, neutral, pride, sadness, surprise). The fields were equally distributed over the screen (2 columns, 5 rows) in alphabetical order. This approach was chosen to facilitate responding by avoiding having to search for the intended answer and bias response times. The mouse was chosen as response input device. The task was created including 10 practice trials of facial emotion recognition in addition to the 360 experimental trials to allow participants to familiarise themselves with the task. This Mediterranean actor was only used in the practice trials and did not appear again anywhere in the experiment to avoid a familiarisation advantage. The experiment was set up to record the correctness per trial (1 = correct, 0 = incorrect). Note: There was no correct or incorrect in this experiment, only a match or non-match with the intended expression defined by the original ADFES (van der Schalk et al., 2011). The display screen resolution for the whole experiment was set to 1280 x 1024 on a 19'' monitor, which allowed the face stimuli to appear in real-size as occurring in face to face social interactions due to the video size of 1024 x 786.

Materials.

The Self-Assessment Manikins (SAM; Bradley & Lang, 1994) of valence and arousal were applied to assess the participants' affective states in both pilot studies presented here. A short neutral video clip was used as part of the experiment for participants to watch, which was validated as being neutral by an independent sample of participants (see Appendix A). A neutral affective state can assure the exclusion of congruency effects in facial emotion recognition experiments where emotions are better recognised when they match the current affective state (Schmid & Schmid Mast, 2010), or diminished recognition due to negative affective states (Chepenik, Cornew, & Farah, 2007), or diminished recognition of certain emotions in a positive affective state (Blairy, Herrera, & Hess, 1999).

The Autism-Quotient-Short (AQ-S; Hoekstra et al., 2011) was used to screen for an ASD in Pilot Study 2. The AQ-S consists of 28 items scored on a 4-point Likert scale, with answer categories "1 = definitely agree"; "2 = slightly agree"; "3 = slightly disagree" and "4 = definitely disagree" resulting in a sum score of minimum 28 and maximum 112. The scoring is reversed for the items where 'agree' is indicative of autism-like traits. The suggested cut-off in clinical screening of autistic traits is a sum score of 65 (Hoekstra et al., 2011).

Procedure*.*

The testing sessions were conducted in a Psychology laboratory at the University of Bath with the experimenter present throughout. Participants were informed about the purpose of the study and written informed consent was obtained prior to participation. Participants were tested individually for Pilot Study 1; two participants were always tested simultaneously within Pilot Study 2, which is why they wore headphones throughout the experiment to avoid disturbances. The experiment started with participants completing the SAM affective rating scales on valence and arousal (Bradley & Lang, 1994). This was followed by presentation of the neutral clip participants watched for 4:18 minutes to allow for participants to settle in and then participants completed the SAM affective rating scales once again. Participants were shown the answer screen to 1) familiarise them with the possible answer categories and 2) to check if they were familiar with the emotional categories (terms), since many international students took part whose first language was not English. Participants were asked to read all the possible answer choices and if a term was unclear, a definition was provided from the Oxford English Dictionary (Simpson & Weiner, 1989). This was then followed by the facial emotion recognition part, with initial practice trials for all 10 possible emotional expressions before the 360 experimental trials (12 actors x 10 expressions x 3 intensities) started.

Each trial started with a fixation cross centred on a blank screen in form of a plus sign (+) presented for 500ms. The fixation cross aims to prepare the participant for the upcoming stimulus and to direct their focus of attention on the screen where the stimulus will appear. The video stimulus was presented for 1040ms following the fixation cross. The experiment was programmed in a way that the mouse cursor disappeared when not needed and only appeared for the emotion labelling. This way, the mouse cursor could not serve as distraction from the video on the screen. A blank screen followed the stimulus presented for 500ms to allow the afterimage, which is automatically created by the video in the observer, to disappear before the answer screen appeared. The ten answer choices were: anger, contempt, disgust, embarrassment, fear, happiness, neutral, pride, sadness, and surprise. As a consequence, chance level for correct responding was 10% (1 out of 10). Participants had to click on one of the emotion terms with the mouse cursor. The experiment was programmed in a way that the 10 answer fields on the answer screen were defined as the only valid fields on the screen. That is, if participants failed to hit one of the valid answer fields, the answer slide re-appeared. This process is not visible to the human eye, but gives the participant the chance to make a valid choice and avoids missing values to occur. The next trial always started after a response was given. Instruction was to have the hand always on the mouse to allow for immediate responding, which was also instructed. No feedback was provided about the correctness of the response. An infinite response time was chosen to not restrict participants' answer time and thereby avoid missing -values to occur. After the emotion recognition experiment was completed, participants of Pilot Study 2 were asked to fill out the paper-based questionnaire (AQ-S; Hoekstra et al., 2011). Completion of the questionnaire took approximately 10 minutes. Duration for completion of Study 1 was approximately 30 minutes and approximately 45 minutes for Study 2. All participants were fully debriefed after completion of the study.

Data preparation and analysis.

Analyses were conducted using SPSS version 21 (SPSS IBM, Armonk, USA). For Pilot Study 1, item difficulty indices (*p* values) were calculated for each of the 360 videos. Descriptive examination of the *p* values was applied where *p'*s greater .10 and smaller 1.00 were desired. (The item difficulty index is equivalent to the accuracy of response for each video expressed in numbers from 0.00 to 1.00, i.e. 0-100%).

For Pilot Study 2, boxplots were used to check for extreme values in all variables of interest (10 emotion categories and 27 emotion by intensity categories). The SPSS default settings were used for determination of extreme values with values more than three times the interquartile range (*IQR*) constituting extreme values. No extreme values were identified for any of the variables of interest. According to the suggestion of Field (2009) for small sample sizes, skewness and kurtosis values were examined to check for normality in the distribution of the data. A non-normal distribution was defined as standardised skewness and kurtosis values greater +/- 1.96. For the 10 emotion categories normality was assumed, as all skewness and kurtosis values were < +/- 1.96. Two of the 27 emotion by intensity categories were non-normally distributed based on the *z*-values of skewness and kurtosis: contempt at intermediate intensity (skewness = 1.88, *SE* = .85, *z*-skewness = 2.23; kurtosis = 3.59, *SE* = 1.74, *z*-kurtosis = 2.06) and fear at low intensity (skewness = 1.81, *SE* = .85, *z*-skewness = 2.13, kurtosis = 3.55, *SE* = 1.74, *z*-kurtosis = 2.04). Since those two variables were only slightly skewed, one sample *t*-tests were conducted for all variables to investigate whether the recognition rates for each variable of interest were significantly above chance level of responding (10%). Each emotion category was displayed by 12 encoders each at 3 intensity levels, so the test value for the 10 emotion categories was 3.6 (10% of 36 trials); the test value for the emotions at each intensity level (27 categories) was 1.2 (10% of 12 trials). For facilitation of interpretation of the results the means and standard deviations are presented in decimal fractions ranging from .00 to 1.00 equivalent to 0-100% accuracy of response.

Problems with experiment (Pilot Study 2).

There were technical problems with the experiment for Pilot Study 2. The stimulus presentation software used (E-Prime) crashed for 1 participant and 3 participants aborted the experiment due to the interruptions that were caused by a technical fault. That is, for some trials the mouse cursor was not visible when the answer slide was presented and therefore participants could not see where they clicked on with the mouse. This was likely to bias the results, as it is uncertain if the participants in those specific trials would have chosen the correct answer. All participant experienced this problem. However, the number of trials where the mouse cursor was invisible varied between participants. The exact number of flawed trials is unknown; the estimated maximum per person is around 20 times of 360 trials. It is also unknown which trials have been affected. As a result, the retrieved results for this study must be interpreted with care.

Results from Pilot Study 1

The majority of the videos (317 of 360) had a *p* value above .10 (chance level of responding) and below a *p* value of 1.00 (absolute ceiling). As desired, the highest count of videos was found for the *p* values of .80 to .95 (150 videos), reflecting good recognisability. The distribution of the *p* values for the whole stimulus set based on each video is illustrated in Figure 5.

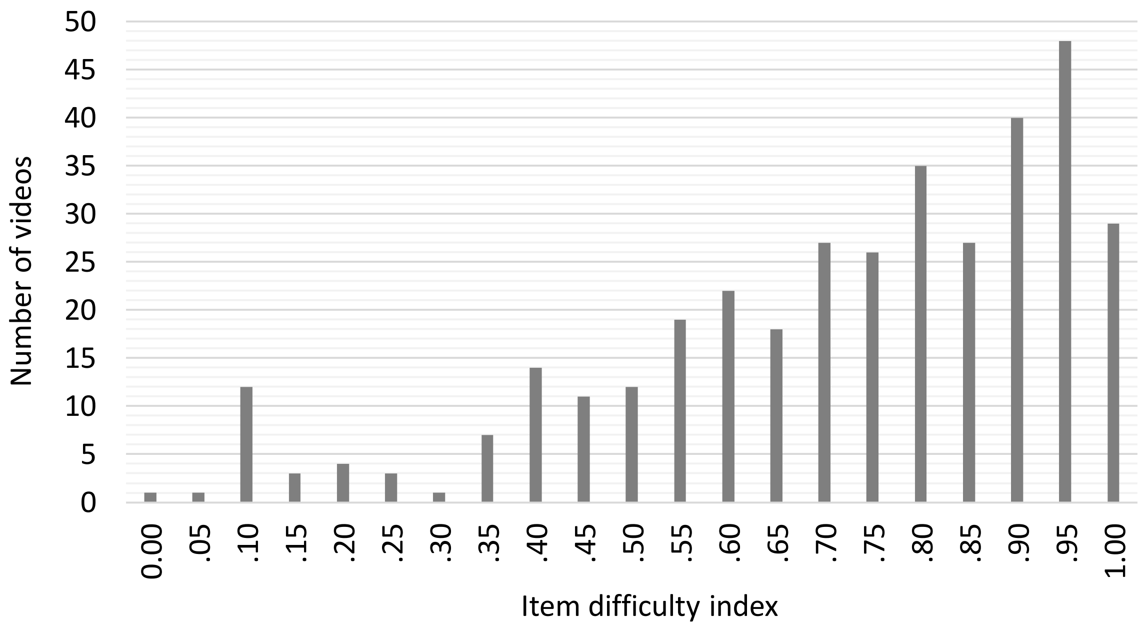


Figure 5. Distribution of item difficulty (*p* value) for the 360 developed videos.

Against the prediction, some videos were found to be recognised at and below a *p* value of .10 and some at 1.00. Fourteen videos (~4% of all videos) were recognised at and below chance level of 10%, of which 5 videos were from the category *contempt* and 6 from *embarrassment*; see Table 1. Visual inspection of these 14 videos suggested that the videos did not provide enough emotional information to recognise the emotions.

Table 1

Video Stimuli with Recognisability at or Below Chance Level of Responding

Emotion	**Intensity**	**Encoder**	***p* value**
contempt	low	F01	.05
contempt	low	F03	.10
contempt	low	F05	.00
contempt	high	F05	.10
contempt	low	M06	.10
fear	low	F01	.10
disgust	low	F03	.10
embarrassment	low	M02	.10
embarrassment	intermediate	M02	.10
embarrassment	low	M08	.10
embarrassment	intermediate	M08	.10
embarrassment	high	M08	.10
embarrassment	low	M11	.10
pride	low	M02	.10

Note. M = male encoder, F = female encoder, *p* value = item difficulty index.

Twenty-nine videos (8%) were recognised by all participants with a *p* value of 1.00. The majority of these videos (14) were *surprise* facial expressions (F01 surprise high intensity; F02 surprise high intensity; F03 surprise high intensity; F04 surprise intermediate intensity; F05 surprise high intensity; F05 surprise intermediate intensity; M02 surprise intermediate intensity; M03 surprise high intensity; M04 surprise high intensity; M04 surprise low intensity; M06 surprise high intensity; M11 surprise low intensity; M11 surprise high intensity; and M12 surprise intermediate intensity). Seven of the videos with a *p* value of 1.00 were expressions of *happiness* (F02 happiness high intensity; F03 happiness intermediate intensity; F05 happiness intermediate intensity; M02 happiness high intensity; M04 happiness high intensity; happiness intermediate intensity; M06 happiness high intensity). Four *neutral* expression videos led to recognition by the whole sample (F02 neutral 1; F02 neutral 2; F04 neutral 1; F05 neutral 3). Two videos expressing *sadness* achieved a *p* value of 1.00 (F01 sadness high intensity; M04 sadness high intensity), as well as one video of *disgust* (F04 disgust high intensity) and one of *anger* (M06 anger high intensity).

Results from Pilot Study 2

AQ-S.

The mean in the sample for the AQ-S (Hoekstra et al., 2011) was 75.80 (SD = 11.10) with a minimum of 56 and maximum sum score of 88. The sample mean exceeding the suggested cut-off of 65 in clinical screening of autistic traits, supporting participants' diagnoses of an ASD, although two individuals did not reach the cut-off score.

Recognition of the 10 emotion categories.

Results of the one sample *t*-tests showed that the following emotion categories yielded recognition rates significantly higher than chance level of responding: neutral (M = .81, SD = .09, $t(5)$ = 18.95, p < .001), happiness (M = .75, SD = .17, $t(5)$ = 9.24, p < .001), surprise (M = .68, SD = .20, $t(5)$ = 7.00, p = .001), sadness (M = .55, SD = .26, $t(5)$ = 4.17, p = .009), disgust (M = .42, SD = .19, $t(5)$ = 4.08, p = .010), fear (M = .34, SD = .22, $t(5)$ = 2.64, p = .046), and embarrassment (M = .43, SD = .25, $t(5)$ = 3.25, p = .023). A trend was found for recognition of anger to be higher than chance level of responding (M = .26, SD = .18, $t(5)$ = 2.28, p = .072). Two categories were not found significantly different from chance level of responding: pride (M = .25, SD = .18, $t(5)$ = 1.86, p = .122), and contempt (M = .12, SD = .17, $t(5)$ = .30, p = .777).

Recognition of the 27 emotion categories.

Results of the one sample *t*-tests showed that for most emotion categories at each intensity level the accuracy of response rates were significantly greater than chance level of responding: surprise low intensity (M = .60, SD = .20, $t(5)$ = 6.09, p = .002), surprise intermediate intensity (M = .75, SD = .20, $t(5)$ = 7.80, p = .001), surprise high intensity (M = .68, SD = .36, $t(5)$ = 4.00, p = .010), happiness low intensity (M = .57, SD = .31, $t(5)$ = 3.77, p =

.013), happiness intermediate intensity (*M* = .85, *SD* = .15, *t*(5) = 11.97, *p* < .001), happiness high intensity (*M* = .85, *SD* = .16, *t*(5) = 11.32, *p* < .001), sadness low intensity (*M* = .51, *SD* = .32, *t*(5) = 3.19, *p* = .024), sadness intermediate intensity (*M* = .49, *SD* = .24, *t*(5) = 3.97, *p* = .011), sadness high intensity (*M* = .65, *SD* = .27, *t*(5) = 4.38, *p* = .007), disgust low intensity (*M* = .39, *SD* = .14, *t*(5) = 5.20, *p* = .003), disgust high intensity (*M* = .47, *SD* = .16, *t*(5) = 5.56, *p* = .003), fear high intensity (*M* = .51, *SD* = .28, *t*(5) = 3.61, *p* = .015), anger high intensity (*M* = .36, *SD* = .23, *t*(5) = 2.74, *p* = .041), embarrassment low intensity (*M* = .36, *SD* = .21, *t*(5) = 3.06, *p* = .028), embarrassment high intensity (*M* = .56, *SD* = .23, *t*(5) = 3.45, *p* = .018). Results also indicated trends for three categories as being recognised higher than chance level of responding: anger intermediate intensity (*M* = .29, *SD* = .19, *t*(5) = 2.50, *p* = .055), disgust intermediate intensity (M = .40, SD = .33, *t*(5) = 2.24, *p* = .075), and embarrassment intermediate intensity (*M* = .36, *SD* = .27, *t*(5) = 2.35, *p* = .066). One-sample *t*-tests showed that 9 categories were not significantly different from chance level of responding: anger low intensity (*M* = .14, *SD* = .17, *t*(5) = .55, *p* = .604), fear low intensity (*M* = .21, *SD* = .24, *t*(5) = 1.11, *p* = .319), fear intermediate intensity (*M* = .29, *SD* = .23, *t*(5) = 2.00, *p* = .101), pride low intensity (*M* = .24, *SD* = .23, *t*(5) = 1.44, *p* = .211), pride intermediate intensity (*M* = .24, *SD* = .20, *t*(5) = 1.67, *p* = .157), pride high intensity (*M* = .26, *SD* = .23, *t*(5) = 1.73, *p* = .144), contempt low intensity (*M* = .07, *SD* = .11, *t*(5) = -.68, *p* = .529), contempt intermediate intensity (*M* = .13, *SD* = .20, *t*(5) =.31, *p* = .767), and contempt high intensity (*M* = .17, *SD* = .23, *t*(5) = .71, *p* = .509).

Discussion

Pilot Study 1 aimed to select a video stimulus set on basis of which videos of facial emotional expressions at varying intensities were developed. It was further aimed to evaluate the developed videos in terms of their recognisability based on their individual item difficulty indexes. The ADFES (van der Schalk et al., 2011) was identified as a suitable video stimulus set for the current research project. Results showed that the majority of the videos (96%) that were edited to varying degrees of intensity of expression were displaying recognisable facial emotional expressions after editing based on *p* levels above the chance level of responding. Against the expectation that all 360 videos would be recognised above chance level of responding but below a *p* level of 1.00, 4% of the videos were recognised below and at chance level of responding and 8% of the videos were recognised by all participants ($p = 1.00$). However, it is common for stimulus sets to contain such items, which is discussed in the following section. The aim of Pilot Study 2 was to investigate whether or not the facial emotion recognition task was feasible for individuals with a diagnosis of high-functioning ASD. Against the expectation results showed that most but not all of the emotion categories and also emotion categories by intensity level yielded accuracy rates significantly above chance level of responding. However, since there were technical problems that affected the recognition rates and therefore the reliability thereof, it can be assumed that the facial emotion recognition task is appropriate for application in a high-functioning autism sample.

By editing the ADFES to videos of varying expression intensity, task difficulty was increased. This is based on the fact that low intensity facial expressions contain less emotional information than high intensity expressions, which is what makes them more ambiguous in nature (Motley & Camden, 1988) and hence harder to identify. That only 4% of all 360 videos were identified with an item difficulty index at and below the set chance level of responding of 10% means that the task difficulty was not overly high. This can be interpreted as a promising outcome for this first evaluation of the developed videos, especially since Bänziger et al. (2009) reported the same percentage of stimuli that were recognised below chance level of responding from a stimulus set they developed and validated but includes high intensity expressions only (MERT). The amount of unrecognised videos from the current study therefore aligns with published facial emotions expression sets.

The few videos that were identified with *p* levels at or below .10, were mainly expressions of complex emotions (contempt, embarrassment, pride) at low intensity of expression. This result is unsurprising, since complex emotions are harder to recognise than basic emotions (e.g. Elfenbein & Ambady, 2002; van der Schalk et al., 2011) and especially hard to recognise at low intensity of expression based on the diminished emotional content available in low intensity expressions. It has been proposed that people are generally less competent at recognising subtle expressions (Matsumoto & Hwang, 2011). This proposition is supported by research from static morphed images of varying intensities which showed that accuracy of response is linearly associated with physical expression intensity (Hess et al., 1997). That is, the lower the intensity of the emotional facial expression, the lower the accuracy rates. It is possible that complex emotions expressed at low intensity are even more ambiguous due to limited available emotional cues and similarity to other emotional expressions. Nonetheless, it is also possible that the videos that with *p* levels of or smaller than .10 were not edited in a way that they contained the necessary emotional information for decoding.

The results of the current study revealed that 40% of the stimuli had *p* values greater than .80. Bänziger et al. (2009) compared five face stimulus sets to each other, the Multimodal Emotion Recognition Test (Bänziger et al., 2009), the Profile of Nonverbal Sensitivity (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), the JACFEE (Matsumoto & Ekman, 1988), the Diagnostic Analysis of Nonverbal Accuracy (Nowicki & Duke, 1994), and the Emotion Recognition Index (K. Scherer & U. Scherer, 2011). They reported that 28-62% of the stimuli in those sets yielded *p* values greater than .80. Thereby the result of the current study falls in between the results achieved from published stimulus sets of facial emotion. The recognisability of the current videos can therefore be considered as satisfying.

It were mostly neutral and high intensity facial expressions of the emotions happiness and surprise that yielded *p* values of 1.00, meaning they got recognised by all participants. This result is not surprising and difficult to avoid, since happiness and surprise are reported throughout the literature as the emotions leading to highest accuracy rates (e.g. Calvo, Gutiérrez-García, Fernández-Martín, & Nummenmaa, 2014; Tottenham et al., 2009). High intensity expressions could not get re-edited, because making those videos harder to recognise (without manipulating the videos) would mean to lower the displayed intensity, but then the resulting video would not portray a high intensity expression anymore. Hence, it was not an option to re-edit those videos. Elimination of those videos would make the number of videos per emotion category uneven and was therefore undesirable. Even though the videos with *p* values of 1.00 have no discriminative ability, it does have no negative effects of having some such items included in the stimulus set.

The emotion categories that did not lead to recognition significantly greater than chance level of responding in the ASD sample from Pilot Study 2 were the emotions pride and contempt. Pride and contempt belong to the category of complex emotions and are harder to recognise than basic emotions, also for individuals without autism (e.g. Elfenbein & Ambady, 2002; van der Schalk et al., 2011). The low contempt recognition rate in the current study is in line with the published literature on English-speaking samples (e.g. Matsumoto & Ekman, 2004; Tracy & Robins, 2008; Wagner, 2000). The low contempt recognition does hence not necessarily imply that the videos are of bad quality. It is possible that the prototypical display of contempt was captured in the videos, but a different label (e.g. disapproval) would represent better the facial expression. It is therefore not surprising that the ASD sample had difficulty recognising these emotions in the conducted experiment. At the lower expression intensity levels, fear and anger were also not recognised significantly above chance level of responding. The recognition of anger and fear is often reported in the literature on facial emotion recognition in ASD to be impaired (e.g. Ashwin et al., 2006; Gioia & Brosgole, 1988; Howard et al., 2000; Pelphrey et al., 2002). This could explain the recognition rates not significantly higher than chance in the current study. However, the results from Pilot Study 2 must be interpreted with care. Due to the technical difficulties that prevented the participants to choose a label intentionally for some trials, the reliability of the results of the current study is diminished. It must be assumed that the accuracy of response rates would have been slightly higher, had there not been the technical fault. The very small sample size of six further decreases the reliability of the results. The small sample size however is a reflection of the difficulty to recruit from the ASD populations.

For the participants from Pilot Study 2 it was not fully assured that the participants had an ASD. The gold standard in confirming an ASD diagnosis is the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012), which takes 40-60 minutes to administer and has to be carried out by a certified examiner. Due to the lack of certification (a very intensive and expensive procedure), a lack of financial resources to outsource ADOS testing, and the restricted time available for testing (as part of the Autism Summer School), the ASD participants within this research project were not ADOS tested. However, to screen for a potential ASD, the AQ-S was applied. Whereas the sample mean supported diagnoses of an ASD, for two individuals the suggested cut-off for a potential ASD was not reached. Since the AQ-S is the short version of the AQ (Baron-Cohen et al., 2001), it is possible that the AQ-S is less suitable for the assessment of autism-traits than the long version. The developers of the AQ-S suggested that when time is not constraint, the long version of the AQ should be applied, especially when communication difficulties are the main interest of research

(Hoekstra et al., 2011). For the following study within this dissertation on an ASD sample, the full version will hence be applied instead of the AQ-S.

Since Pilot Study 1 revealed that most of the videos from the ADFES (van der Schalk et al., 2011) that were edited to varying levels of intensity fell in the desired range of recognition between chance level of responding and maximum recognition, the videos that were recognised by all participants remained included in the stimulus set. However, it was decided to attempt to create new sequences for the 14 videos that were recognised at and below the chance level of responding. The undertaken amendments are presented in the following Chapter 5. Pilot Study 2 showed that most categories included in the stimulus set were correctly recognised significantly above chance level of responding by a sample of individuals with high-functioning ASD, which means that the experiment can be considered feasible to be undertaken by individuals who struggle with facial emotion recognition. Therefore, the adaptations of the ADFES videos seemed suitable for further application within this research project. The stimulus set including the newly created videos is from here on called the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV). In the stimuli development section of this chapter it was described that the expression intensity levels were created based on my subjective judgements. However, to allow for application of the ADFES-BIV within this research project, the created intensity levels need to be confirmed on a sample of independent judges. The respective study is presented in the following chapter.

CHAPTER 5—Confirmation of the created ADFES-BIV intensity levels

Introduction

Varying levels of expression intensity can be created in a highly standardised manner when using the morphing technique, because anatomical reference points in the face and difference vectors between these points need to get defined (e.g. the position of the mouth corner in neutral and when smiling at full intensity) and percentages of expression intensity represent the various points on those vectors. For example, when specifying 10% intensity the morphed image will represent the point at 10% of the full vector scale from neutral (see Wild, Erb, & Bartels, 2001). This mathematical approach to determine intensity of expression is not possible to take when using video-recordings instead of such computer-generated pictures. For the creation of the three intensity levels of the ADFES-BIV the frame representing either low, intermediate, or high intensity of the emotional facial expression for each video was a subjectively determined (see Chapter 4). It was therefore necessary to confirm the created intensity levels of the ADFES-BIV on a sample of independent judges.

Aims and hypotheses

Aim of the study presented in this chapter was to confirm the intensity levels of the ADFES-BIV as distinct categories based on the judgements of perceived intensity of facial expression on a student sample.

It was hypothesised that:

(1) the high intensity facial expressions would be judged as more intense than the intermediate intensity expressions and that the intermediate intensity facial expressions would be judged as more intense than the low intensity expressions.

(2) the pattern outlined in hypothesis 1 would be present for all nine emotions included in the ADFES-BIV.

Method

Participants.

Thirty-five individuals (15 females, 20 males) were recruited from the University of Bath community through advertisement on campus in form of posters, word of mouth, and through the web-noticeboard. Additionally, the Research Participation Scheme was used to recruit first year Psychology Undergraduate students who are required to undertake participation hours as part of the Psychology degree and receive course credit in return. Due to a technical fault, data were only available for 34 participants (-1 male). Participants were PhD students (*n* = 9), Undergraduate students (*n* = 21), and staff (*n* = 4) and came from diverse departments including Psychology (*n* = 27), Health (*n* = 1), Mechanical Engineering (*n* = 4), Social Policy and Science (*n* = 2), and Physics (*n* = 1). All participants had normal or corrected-to-normal vision to assure the ability to perceive the presented emotion cues. Participants were not monetarily compensated, but Psychology students gained course credit for participation. The age ranged from 18 to 41 years (*M* = 23.7, *SD* = 6.1). One person indicated having a diagnosis of an Anxiety Disorder. Since the data for this participant was comparable to the other participants in terms of their intensity judgements, their data remained included in the analysis. Ethical approval for this study was given by the University of Bath Psychology Ethics Committee via Chairs Action (Ref 13-044).

Stimuli.

Based on the results from the pilot stimuli evaluation (Pilot Study 1 in Chapter 4), some amendments were undertaken. The 14 identified videos with accuracy rates at or below chance level (F01 contempt low intensity, F01 fear low intensity, F03 contempt low intensity, F03 disgust low intensity, F05 contempt low intensity and high intensity, M02 embarrassment low intensity and intermediate intensity, M02 pride low intensity, M06 contempt low intensity, M08 embarrassment low intensity, intermediate intensity, and high intensity, M11 embarrassment low intensity) were visually inspected. This inspection revealed that most of the videos were cut too early, not providing enough information for facial expression decoding. Hence, those videos were edited again following the procedure described in Chapter 4. For some videos the low intensity video was dismissed and the intermediate intensity video classified as low intensity instead; Table 2 shows the amendments undertaken.

Table 2

Specifications of the Undertaken Amendments of the Videos that were Recognised at and Below Chance Level of Responding

Encoder/emotion	**Amendments**
F01 contempt	new low intensity video created
F01 fear	new high intensity video created, low intensity video deleted, intermediate became low intensity video, high intensity video became intermediate
F03 contempt	intermediate became low intensity video, new intermediate and high intensity video created
F03 disgust	intermediate became low intensity video, new intermediate intensity video created, high intensity video kept
F05 contempt	new low intensity video created, intermediate intensity video kept, new high intensity video created
M02 embarrassment	intermediate became low intensity video, new intermediate intensity video created, high intensity video kept
M02 pride	new low, intermediate, and high intensity video created
M06 contempt	new high intensity video created, intermediate intensity video became low, high became intermediate
M08 embarrassment	new low, intermediate, and high intensity video created
M11 embarrassment	new low, intermediate, and high intensity video created

Note. F = female encoder. M = male encoder.

Intensity judgement task.

The 360 ADFES-BIV videos, including the re-edited videos, were used as stimuli for the intensity judgement task. Three of the videos (one for each intensity level) were used for practice purposes before the task and hence appeared twice. The intensity judgements were made with a slide bar on a visual analogue scale ranging from very low to very high and medium in between. The slide bar was operated with the computer mouse. E-Prime was coded that way that percentages were recorded with the left end of the scale representing 0% intensity and the right end of the scale 100% intensity of expression.

Procedure.

Participants were informed about the specifics and purpose of the study and gave written informed consent prior to participation. One to four participants were tested simultaneously. All participants were guided through the three practise trials by the experimenter at the same time. Instruction was to judge the intensity of the facial emotional expression and to make use of the whole scale for the judgements. It was further instructed to judge each video on its own and not to compare between emotions, since some emotions are generally perceived as more intense than others. Participation in the study took approximately 30min. After completion of the experiment participants were fully debriefed and thanked for participation. Psychology students were granted course credit for participation.

Data preparation and analysis.

The percentages of the intensity judgement ratings were the DV and intensity judgements for all categories of the ADFES-BIV (3 intensity levels, 10 emotion categories, 27 emotion by intensity categories) were investigated. The data were inspected for extreme values using boxplots with values more than three times the *IQR* constituting extreme values and the assumption of a normal distribution was checked with Shapiro-Wilk tests. For none of the variables extreme values were identified. The distribution of the variable for the neutral expressions was significantly different from normal ($W(34) = .81$, $p < .001$) with a skew to the right (skewness = 1.24, $SE = .40$, z-skewness = 3.10). Hence, the difference in judgement of intensity of facial expression between the low intensity facial expressions and the neutral expressions was tested using the Wilcoxon matched pairs signed ranks test. The variable for surprise at low intensity was also flagged as significantly different from a normal distribution with a slight skew to the left ($W(34) = .94$, $p = .049$). This slight deviation from normality in the distribution was confirmed by the *z*-skewness value of 2.06 (skewness = -.83, $SE = .40$) and *z*-kurtosis of 2.59 (kurtosis = 2.05, $SE = .79$), which were greater than 1.96. For the remaining 26 variables a normal distribution was assumed (W's(34) > .94, p's > .064). With the vast majority of the variables not being significantly different from a normal distribution and surprise at low intensity only showing slight deviation from normal, a repeated measures ANOVA was conducted with the 9 emotion categories and 3 intensity levels as within-subject factors.

Results

The repeated measures ANOVA indicated a violation of Sphericity for the factors *intensity* ($W(2) = .43$, $p < .001$) as well as *emotion* ($W(35) = .02$, $p < .001$) and the interaction *intensity*emotion* ($W(135) = .00$, $p = .027$). Therefore, Greenhouse-Geisser adjustment of degrees of freedom was applied. There was a significant main effect of *intensity* ($F(1.27,42.05) = 452.70$, $p < .001$, partial $\eta^2 = .932$, power = 1.00). Pairwise comparisons showed that the high intensity facial emotional expressions ($M = 66.19$, $SD = 6.87$) were perceived as being significantly higher in intensity than the intermediate intensity expressions ($M = 55.07$, $SD = 8.16$, $p < .001$) and the intermediate intensity expressions were judged significantly higher in intensity than the low intensity expressions ($M = 42.32$, $SD = 9.39$, $p < .001$); see Figure 6. The difference in judgement of intensity between the neutral expressions ($M = 9.17$, $SD = 9.43$) and the low intensity expressions was indicated as significant based on Wilcoxon matched pairs signed ranks test ($z = -5.07$, exact $p < .001$).

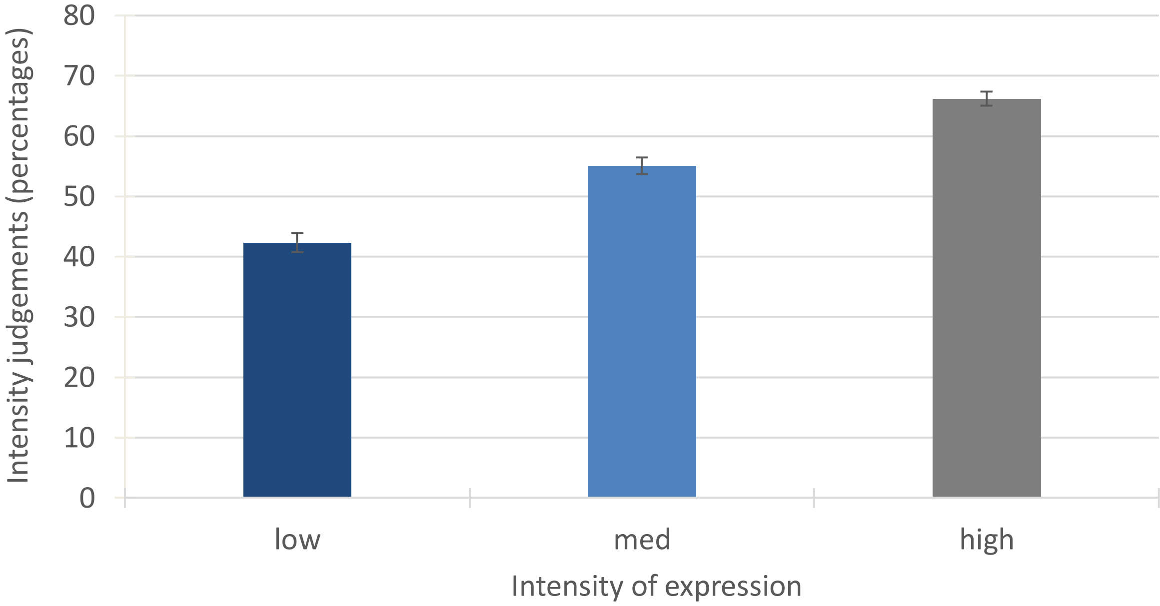


Figure 6. Intensity judgements for the intensity levels of the ADFES-BIV in percentages. Error bars represent standard errors of the means.

The results showed that there was a significant main effect of *emotion* ($F(4.23,139.56) = 151.88$, $p < .001$, partial $\eta^2 = .822$, power = 1.00). Fear was judged as the most intense expression ($M = 72.04$, $SD = 9.13$) followed by surprise ($M = 70.51$, $SD = 9.48$), pride ($M = 64.06$, $SD = 7.71$), disgust ($M = 57.72$, $SD = 10.50$), happiness ($M = 55.59$, $SD = 6.85$), anger ($M = 54.85$, $SD = 10.26$), sadness ($M = 43.88$, $SD = 10.65$), embarrassment ($M = 42.39$, $SD = 12.02$), and contempt ($M = 29.69$, $SD = 11.55$); see Figure 7. Pairwise comparisons showed that anger was significantly different to all emotions (p's < .030) but happiness ($p = .626$). Disgust was also significantly different from all emotions (p's < .030) but happiness ($p = .092$). Fear was significantly different from all other emotions (p's < .025), as so surprise (p's < .025), contempt (p's < .001), and pride (p's < .001). Embarrassment was significantly different from all emotions (p's < .001) but sadness ($p = .300$).

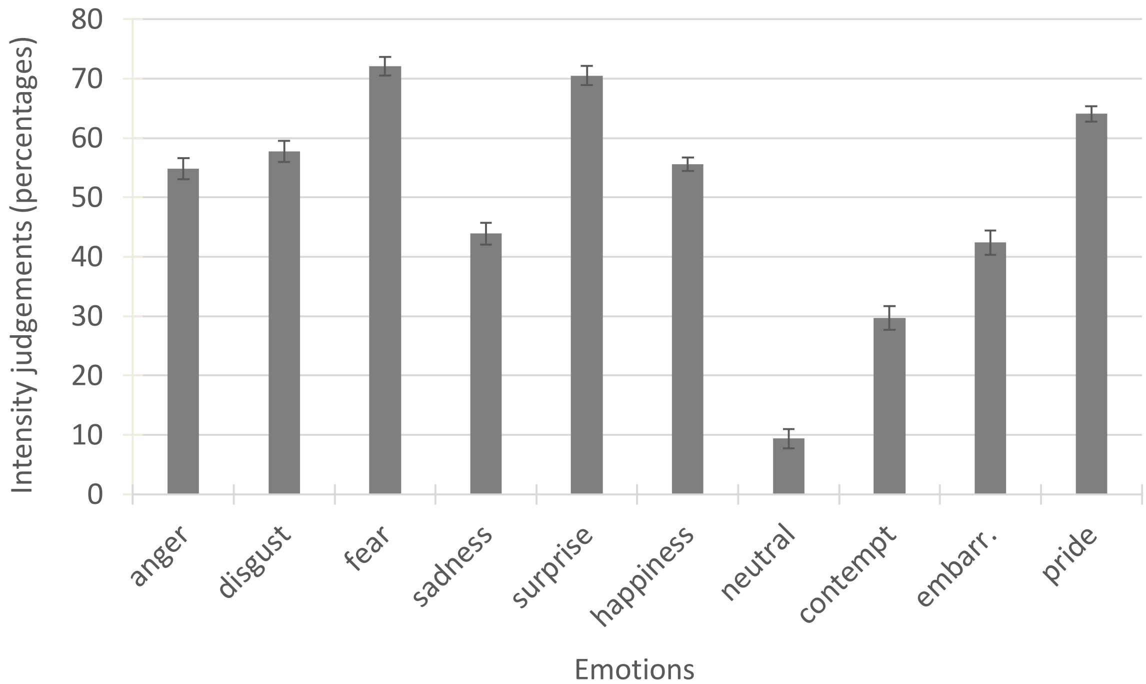


Figure 7. Intensity judgements for the emotion categories and neutral faces of the ADFES-BIV in percentages. Embarr. = embarrassment. Error bars represent standard errors of the means.

The interaction of *emotion*intensity* was also significant ($F(9.50,313.39) = 42.70$, $p < .001$, partial $\eta^2 = .564$, power = 1.00). Pairwise comparisons showed that for each of the emotions the high intensity expressions were judged significantly more intense than the intermediate intensity expressions (p's < .001), which in turn were judged significantly more intense than the low intensity expressions (p's < .001). However, the differences in intensity judgement between the intensity levels differed between the emotions; see Figure 8. Means and *SD*s are given in Table 3.

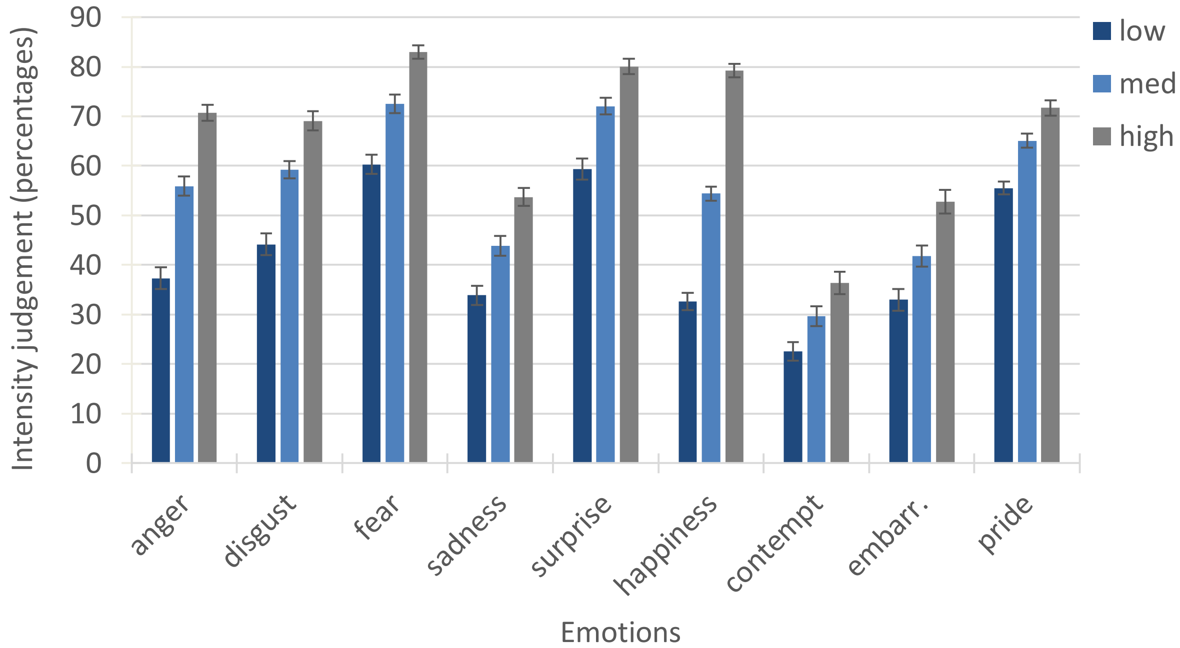


Figure 8. Intensity judgements for the emotion categories of the ADFES-BIV at each intensity level in percentages. Embarr. = embarrassment. Error bars represent standard errors of the means.

Table 3

Means and Standard Deviations of the Intensity Judgements for Each Emotion Category by Intensity Level

Emotion	**Means (Standard Deviations)**		
(*N* = 34)	**low**	**intermediate**	**high**
Anger	37.33 (12.69)	55.91 (11.06)	70.72 (9.53)
Disgust	44.19 (13.11)	59.19 (10.16)	69.06 (11.24)
Fear	60.31 (11.48)	72.51 (10.82)	82.93 (7.85)
Sadness	33.87 (11.61)	43.84 (11.69)	53.7 (10.60)
Surprise	59.31 (12.43)	72.06 (9.95)	80.04 (9.11)
Happiness	32.68 (10.20)	54.42 (8.35)	79.22 (7.74)
Contempt	22.58 (10.89)	29.68 (11.97)	36.35 (13.22)
Embarrassment	33.01 (12.82)	41.82 (12.25)	52.73 (13.83)
Pride	55.54 (7.29)	65.05 (8.24)	71.70 (9.13)

Note. Means (*M*) and standard deviations (*SD*) are presented in percentages.

Discussion

The study presented in this chapter aimed to confirm the three created intensity levels of the ADFES-BIV as overall categories and for each emotion category based on judgements of perceived intensity of expression from independent raters. In line with the expectation, the results showed that the high intensity facial expressions were perceived as significantly more intense as the intermediate intensity expression, which in turn were perceived as significantly more intense than the low intensity expressions, overall and at the level of individual emotions. This pattern is in line with previous reports of increasing intensity ratings with increasing intensity of expression based on morphed images (Hess et al., 1997). Results further demonstrated that the low intensity facial expressions were judged as significantly more intense than neutral faces. Overall, the results show that the intensity levels of the ADFES-BIV are distinct from each other.

Interestingly, the results from the study presented in this chapter revealed that the intensity ratings differed between the emotion categories and the differences in intensity judgements for the intensity levels varied for the individual emotions. Fear and surprise were the emotions that were rated as most intense basic emotions and pride the most intense of the complex emotions. Contempt was judged as the least intense complex emotion and overall and sadness was the least intense basic emotion. It is possible that the amount of activated facial features within an emotional facial expression determine the intensity it conveys, but it is also possible that facial features that allow for greater movements are perceived as more intense. For example, a wide open mouth is a larger movement than a pulled down mouth corners and might hence be perceived as more intense. However, it is also plausible that these intensity judgements reflect the underlying intensity of arousal. In fact, Widen and Naab (2012) found that facial emotional expressions including an open mouth were rated higher in arousal than facial expressions showing the mouth closed. The found differences in intensity judgement from the current study might simply represent arousal, as fear is generally judged as a high arousal emotion and sadness as a low arousal emotion (see Widen and Naab, 2012). However, judgements of intensity of expression might be driven by the amount of emotional information available in the face and the arousal level attributed with the emotion.

The results from the study presented here show that a video stimulus set of facial emotion was created that includes varying degrees of expression intensity. Even though in Chapter 4 the videos were evaluated on item level, which showed that 96% of the videos were recognisable, this does not constitute a sufficient validation. A validation of the categories included in the ADFES-BIV is necessary. Further validation is especially necessary since the ADFES-BIV includes re-edited videos that were not yet tested for recognisability. The validation of the ADFES-BIV on a larger sample is presented in the following chapter.

CHAPTER 6—Validation of the ADFES-BIV on a general population sample

Introduction

Video stimuli of facial emotion were selected, edited to three levels of expression intensity, and evaluated within a pilot study on a small sample in Chapter 4. Even though the results showed generally good recognisability of the developed videos (with only 4% of the 360 videos that were not recognised above chance level of responding), a more comprehensive validation of the stimulus set was necessary as a next step. Such a validation comprises of validation of all categories included in the stimulus set. Accuracy of response rates constitute the main characteristic of evaluation to validate stimuli for application within facial emotion recognition research and recognition rates significantly greater than chance level of responding are generally required for validation.

The results from Chapter 4 showed that some emotions are more difficult to recognise than others. For example, the most videos that were not well recognised were such portraying complex emotions of contempt and pride and most of the not well recognised videos were low intensity facial emotional expressions. The recognition rates reported by the developers of the ADFES (van der Schalk et al., 2011) also demonstrate that emotions differ in their recognisability. The non-negative emotions surprise and happiness are generally reported as well recognisable (e.g. Tottenham et al., 2009) and fear is often reported as the emotion achieving the lowest accuracy of response rates from the basic emotions (e.g. Calvo & Lundqvist, 2008; Kirouac & Dore, 1985). That the intensity of facial expression also influences recognisability was demonstrated by Hess et al. (1997) as they found accuracy of response to increase linearly with increasing physical expression intensity. This finding implies that the stimuli that are easier to recognise lead to higher accuracy of response. In addition, participants are faster at recognising the easier to recognise emotion happiness than the more difficult to recognise emotion anger based on response latencies (Kestenbaum & Nelson, 1992). It can be concluded that the recognisability of an emotional facial expression measured on accuracy of response and response latencies depends on the emotion category and the intensity of expression. However, not much is known about how the intensity of a facial emotional expression influences the recognisability of specific individual emotions, especially not from video stimuli.

Next to validating a stimulus set based on recognition rates of the stimuli, the sensitivity of the stimuli to differentiate between individuals' performance at facial emotion recognition can be applied as a measure of quality. Stimuli of high sensitivity are thereby expected to reveal group differences that are being investigated but are based on a small effect size. A known group difference in facial emotion recognition that is thought to be based on such a small effect is sex differences in facial emotion recognition. Precisely, females generally outperform males in facial emotion recognition, i.e. achieve higher accuracy rates (meta-analysis by J. A. Hall 1978, 1990; literature review by Kret & De Gelder, 2012); as has been mentioned in Chapter 1. It has to be noted that some studies did not find a female advantage when investigating sex differences in facial emotion recognition (e.g. Grimshaw, Bulman-Fleming, & Ngo, 2004; Rahman, Wilson, & Abrahams, 2004). However, it is possible that the supposed small effect size is underlying non-significant sex differences in facial emotion recognition is based on a small effect. For example, JK Hall et al. (2010) reported an effect size of partial η^2 = .13. For effect sizes that small, next to a sufficient sample size, a sensitive task is needed to retrieve significant results. With a sufficient sample size and a sensitive task the female advantage should reflect in the results. One way of increasing task sensitivity for a facial emotion recognition task is the inclusion of varying levels of intensity of expression. However, studies investigating sex differences in recognition of facial emotional expression from varying intensities are sparse.

There are at least two studies of facial emotion recognition where the intensity of expression was varied. Hoffmann, Kessler, Eppel, Rukavina, and Traue (2010) conducted two experiments investigating the recognition of the six basic emotions and contempt at varying intensities of facial emotional expression in males and females. The varying intensities were created using the morphing technique, but presented to the participants as static images. Whereas in the first experiment full-blown expressions (100%) and such of medium intensity (50%) were used, the second experiment applied a range of intensity from 40% to 100% (in 10% steps). In both experiments, Hoffmann et al. (2010) found females to outperform males only for recognition of subtler facial expressions averaged over all emotions but not high intensity expressions (i.e. 80-100%). The results indicate that sex differences only become apparent when difficulty to recognise the emotion is high enough, which is also an indicator of a sensitive task. At the level of individual emotions the emerged differences were inconsistent between intensity levels and experiments. In contrast to Hoffmann et al. (2010), JK Hall et al. (2010) found that there are no emotion-specific differences between males and females regarding the recognition of emotions. JK Hall et al. (2010) investigated sex differences in regards to accuracy of response and response latencies of facial emotion recognition of the six basic emotions using intensity variations ranging from 30% to 100%, which were also based on morphed static images. It has to be noted though that they averaged their data over the intensities for each emotion. An overall female advantage in accuracy of response and response times was reported, but the interaction of emotion and sex was not significant. To be able to reveal potential sex differences for individual emotions, it might be necessary to investigate the emotions on each of the intensity levels as done by Hoffmann et al. (2010). Besides the limitations of having applied static stimuli and having averaged the accuracy rates across intensity levels, the study conducted by JK Hall et al. (2010) has the advantage of including response latencies as a DV. It has to be noted that response time is an aspect often neglected in facial emotion recognition research, despite its higher sensitivity to reveal deficits/differences in processing of facial emotional expressions compared to accuracy where ceiling effects often hamper group comparisons (Hampson, Van Anders, & Mullin, 2006). However, there are further reports of females correctly recognising facial emotional expressions faster than males (e.g. JK Hall et al., 2010; Hampson et al., 2006; Lee, Krabbendam, & White, 2013; Rahman et al., 2004).

The female advantage in facial emotion recognition that is often reported leads to the question whether the female advantage is biologically hard-wired or a results of different socialisation processes for males and females, which is still a question that remains to be answered (McClure, 2000). Several theories have been postulated to explain the female advantage, for example the primary caretaker hypothesis (Babchuk, Hames, & Thompson, 1985). The primary caretaker hypothesis is further divided into the attachment promotion hypothesis and the fitness threat hypothesis (Babchuk et al., 1985) and some support has been found for both hypotheses (Hampson et al., 2006). According to the primary caretaker hypothesis, females' greater emotion processing ability is innate to assure proper caretaking of the offspring. The attachment promotion hypothesis (Babchuk et al., 1985) predicts that females can better recognise all emotional facial expressions than males based on the necessity to be responsive to the child so the child develops a secure attachment style. The fitness threat hypothesis (Babchuk et al., 1985) however predicts that females are only better than males at recognising negative emotions from faces, to keep the child safe. Females' greater emotion processing ability is evidenced by the results of JA Hall and Matsumoto (2004) who presented facial expression with exposure times only allowing sub-conscious processing and came to the conclusion that females perceive emotions more in a gestalt fashion than males, making fast and automatic recognition possible. Donges, Kersting, and Suslow (2012) found females to outperform males in perception and responding to positive facial emotion at an automatic processing level and concluded that this perceptual sensitivity underlies the female advantage. These results offer an explanation for why fast and automatic processing of facial emotion would be beneficial for females, as such processing provides a solid basis for apt caretaking of offspring. If the female advantage in facial emotion recognition is based on an innate primary caretaker role, females should be able to detect emotions from faces fast and also be able to detect emotions already from subtle expressions to address to the needs of the child consistent with the reported findings from JK Hall et al. (2010) and Hoffmann et al. (2010).

With a very limited number of studies that included varying intensities of facial emotional expression and their mixed findings, there is a need for investigations including varying intensities. Furthermore, with most studies being based on static images in general, investigations applying dynamic stimuli of varying intensities of facial expression are needed. In real-life social interactions, next to accurate recognition, timely responding is required. Both aspects of facial emotion recognition contribute to functioning social interactions, which makes response time equally important as accuracy (De Sonneville et al., 2002). As a result, investigations including both, accuracy of response and response latencies, are needed. Therefore and for the purpose of further validation of the ADFES-BIV, processing of facial emotion in males and females based on accuracy of response rates and response latencies were investigated in this study on basis of video stimuli containing varying intensities of expression.

Aims and hypotheses

The aims of the current study were to:

(1) *validate* ADFES-BIV, which included testing recognition rates of the different emotion and intensity categories and individual videos against chance level of responding and comparing emotion and intensity categories between themselves. Accuracy of response as well as response latencies were examined to validate the intensity levels. It was aimed to explore how intensity of expressions influences recognition of the individual emotions from the ADFES-BIV.

(2) test the ADFES-BIV's *sensitivity* to reveal group differences in facial emotion recognition ability in the general population that are thought to exist by comparing recognition rates and response latencies of males versus females on all categories of the ADFES-BIV.

Hypotheses to aim 1: Validation of the ADFES-BIV.

(1.1) It was expected that the high intensity level of expressions would yield higher recognition rates and faster responses than the intermediate intensity expressions, which in turn would result in higher recognition rates and faster responses than the low intensity expressions.

(1.2) It was further expected that the accuracy of response rates for the 10 emotion categories of the ADFES-BIV would be greater than chance level of responding.

(1.3) It was hypothesised that there would be differences in recognition rates between emotion categories with the negative basic emotions (fear, anger, disgust, and sadness) having lower recognition rates compared to non-negative emotions (happiness and surprise), while the more complex emotions (embarrassment, price, and contempt) would have the lowest recognition rates.

(1.4) It was expected that the 27 emotion categories at each intensity level would be recognised with an accuracy greater than chance level of responding.

(1.5) It was hypothesised that the intensity of facial expression would influence the recognition rates of the specific emotions to varying degrees. That is, emotions that are generally easy to recognise (e.g. happiness, surprise) would yield in high recognition rates across intensity levels, whereas emotions that are generally harder to recognise (e.g. fear, contempt) would yield in recognition rates considerably lower at low intensity compared to high intensity of expression.

(1.6) Additionally, it was expected that more than 96% of the 360 videos of the ADFES-BIV would be recognised above chance level of responding.

Hypothesis to aim 2: Sensitivity test.

(2) It was expected to find the female advantage in facial emotion recognition based on accuracy rates and response latencies, which would be more prominent for the more difficult parts of the task, i.e. for subtler facial expressions and emotion categories that are more difficult to recognise, i.e. complex emotions and negative basic emotions.

Method

Participants.

Student and staff participants were recruited from the University of Bath through advertisement on campus in form of posters, word of mouth, and through the web-noticeboard. Additionally, the Research Participation Scheme was used to recruit first year Psychology Undergraduate students who are required to undertake participation hours as part of the Psychology degree and receive course credit in return. The sample of the current study consisted of 41 male and 51 female participants (N = 92), aged between 18 and 45 (M = 23.3, SD = 5.7). Participants were mainly enrolled in an Undergraduates program (n = 54), but also Postgraduate students (Masters, n = 8; PhD, n = 22) and staff (n = 8) participated. Participants came from diverse departments with the majority from Psychology (n = 35) and Biology (n = 11). The sample was culturally diverse, with the majority British (n = 55), followed by Asian (n = 15), Southern European (n = 8), Southern American (n = 6), Northern

European (n = 6), and Northern American (n = 2). All participants were assumed to be fluent in English, based on it being one of the entry requirements to study at the University of Bath. All participants had normal or corrected-to-normal vision to assure the ability to perceive the presented emotion cues. None of the participants reported a clinical diagnosis of a mental disorder. Ethical approval for this study was given by the University of Bath Psychology Ethics Committee (Ref 13-137).

Stimuli.

The ADFES-BIV videos as used in the study presented in Chapter 5 were used within this validation study.

Procedure and materials.

The testing sessions were conducted in a laboratory at the University of Bath with the experimenter present. After written consent was obtained participants undertook the computer task (~30min). Participants were either tested individually or in a group with a maximum of four participants simultaneously, following the same procedure as laid out in Chapter 4. In the case of a group testing, participants wore headphones throughout the experiment to diminish disturbances. The facial emotion recognition task presented in Chapter 4 was applied in an identical manner starting with an affective rating using the non-verbal Self-Assessment Manikin of valence and arousal (SAM; Bradley & Lang, 1994) before and after the neutral clip (fertilisers; Appendix A). The neutral clip aimed to settle the participants for the experiment in case any strong and distracting feelings might have been present, especially since positive and negative affective states can alter emotion perception (Niedenthal et al., 2001). Ten practice trials were conducted before the 360 experimental trials. Participants were instructed to imagine actually interacting with the encoders and respond intuitively, making immediate responses necessary. Since there were 10 answer choices for each trial, the chance level of response was 10%. The experiment was set up to record the correctness per trial (1 = correct, 0 = incorrect), and response times (in ms). After completion of the study participants were fully debriefed and either paid £5 for participation or granted course credit.

Dependent variables.

DV 1: *Accuracy of response* reflects the percentages of correct responses out of a number of given trials, the number of which varied at each level. The accuracy rates for each of the *10 emotion categories* were based on 36 trials, since 12 encoders expressed each emotion at three intensity levels (3 x 12 = 36 = 100%). The accuracy rates for the *3 intensity levels* were each based on 108 trials, since 9 emotional expressions were displayed by 12 encoders at each of the three intensity levels (12 x 9 = 108 = 100%). Only nine emotions were included at this level, since the category neutral does not have any intensity levels. The accuracy rates for each of the *27 emotion categories by intensity level* were based on 12 trials, since 12 encoders displayed each emotion at each of the three intensity levels (12 = 100%). For facilitation of interpretation of the results the means and standard deviations are presented in decimal fractions ranging from .00 to 1.00 equivalent to 0-100% accuracy of response.

DV 2: *Response time* refers to the time participants took to respond from the moment the answer screen was presented until the participant clicked the mouse on their answer choice. This is different from a reaction time where participants are instructed to answer as fast as they can. Mean response times were computed for *the same categories* as laid out for DV 1, i.e. 3 intensity levels, 10 emotion categories, 27 emotion by intensity categories. Only trials with correct responses were used in this analyses. Response times were measured in ms.

Data preparation and analysis.

For testing of the outlined hypotheses different types of data preparations and analyses were necessary. Therefore, data preparations and data analyses are presented in the following separated by hypotheses.

Data preparation and analysis *for Hypothesis 1.2.*

Before the accuracy rates of the 10 emotion categories could be tested against chance level of responding, participants' means of accuracy of response for those categories were checked for extreme values using boxplots defined by > 3 times the *IQR.* The distribution of each of the 10 emotion categories was checked for normality using Shapiro-Wilk tests. The Shapiro-Wilk test was suggested as best option by Thode (2002) for testing normality due to its good power. Four extreme values were identified: neutral (participant 35 and 66), sadness (participant 73), and surprise (participant 33). For all these extreme values the mean recognition scores were lower than for the rest of the sample on the respective category. Since these extreme values each belonged to a different participant, they constitute univariate outliers. Since it was the aim of this study to validate the stimuli, excluding data points or changing the scores would increase accuracy rates and was hence refrained from. Data transformations, which can help minimising extreme values, were also inapplicable. This is because the direction of the skews varied between the variables with some left- and some right-skewed. However, the varying types of available transformations are each applicable for either left- or right-skewed data. Additionally, the same type of transformation would have to get applied to all variables (Field, 2009). Therefore, when some variables are skewed to the left and others to the right, transformations are inapplicable. Shapiro-Wilk tests showed that the 10 emotion categories were all significantly different from a normal distribution (W's(92) < .95, p's < .05). Therefore, Wilcoxon signed-rank tests were conducted to compare the medians of the 10 emotion variables to chance level of responding, which was 10% based on 10 emotional category choices. Since multiple comparisons were conducted, a Bonferroni-corrected p-value = .005 (p = .05/10) was applied.

Data preparation and analysis *for Hypothesis 1.4.*

Before the 27 accuracy categories (emotion by intensity) could be tested against chance level of responding, participants' means of accuracy of response for each of those categories were inspected for extreme values using boxplots. Across the 27 categories investigated, seven extreme values with means lower than the rest of the sample were identified by inspecting boxplots: anger intermediate intensity two extreme values (participant 16 and 24), embarrassment high intensity (participant 12), happiness intermediate intensity (participant 80), happiness high intensity (participant 73), surprise intermediate intensity (participant 33), and surprise high intensity (participant 21). Since these extreme values each belonged to a different participant, they constitute univariate outliers, which emerged based on the very small ranges of accuracies for the identified variables. No data was excluded. Since it was the aim to validate the stimuli and the identified extreme values all occurred at the lower end of the *IQR*, excluding data points or changing the scores would increase accuracy rates and was hence refrained from. Data transformations, which can help minimising extreme values, were also inapplicable, explained above. The distribution of each of the 27 categories was checked for normality using Shapiro-Wilk tests. Of the 27 categories for the accuracies of the emotions at each intensity level, 26 variables were significantly different from a normal distribution (W's(92) < .96, p's < .05). Since the direction of the skews varied between the variables with some left- and some right-skewed, transformations could not be performed. Based on the presence of extreme values and violation of normality, non-parametric Wilcoxon signed-rank tests were conducted. The medians of all 27 variables were compared to chance level, which was 10%. Since multiple comparisons were conducted a Bonferroni-corrected p-value = .002 (p = .05/27) was applied.

Data preparation and analysis *for Hypothesis 1.6.*

To test whether each individual video in the stimulus set was perceived accurately above chance level of responding by the sample, the accuracy rates of each individual video of the ADFES-BIV were compared to chance level of responding using the Binomial test. This test was chosen because it is suitable for data where there is no predictor and the outcome variable is dichotomous (as 0 and 1 were coded for correct vs incorrect labelling). Since multiple comparisons were conducted a Bonferroni-corrected p-value = .000139 (p = .05/360) was applied.

Data preparation and analysis *for hypotheses 1.1, 1.3, 1.5, and 2.*

The variations between emotions regarding their recognition rates and response times (hypotheses 1.3), the influence of intensity of expression on the individual emotions (hypothesis 1.5), and the variations between intensity levels regarding their recognition rates and response times (hypothesis 1.1) were examined within the analysis carried out to test for sex differences in accuracy rates and response times (hypothesis 2).

It has to be noted that since the emotion category ‘neutral’ does not include different levels of intensities, it was not included in the analysis. However, males and females were compared on the recognition of neutral to test for a bias in the perception of facial expression. Since the variable ‘neutral’ was significantly non-normal for males ($W(41) = .81$, $p < .001$) and females ($W(51) = .69$, $p < .001$) as indicated by Shapiro-Wilk tests, and although homogeneity of variances was given ($F(90,1) = .60$, $p = .440$), Mann-Whitney *U* test were conducted to compare males and females on their accuracy and response latencies in response to neutral faces. Mann Whitney *U*-tests showed no significant difference in accuracy of response ($U = 939.00$, $z = -.85$, exact $p = .400$) or response latencies ($U = 848.00$, $z = -1.55$, exact $p = .122$). These results suggest that males and females do not differ in face processing on a general level based on processing of neutral faces.

One potential statistical approach on the data would be a repeated measures ANOVA, however, it was decided to utilise generalised linear mixed modelling (GLMM) to test the current data. This is because the ANOVA assumptions of normality and homogeneity of variances were not met for most of the variables included in this study, and the group sizes were unequal. The latter issue makes the *F* statistic unstable when assumptions are violated (Field, 2009). Rather than trying to make the data fit the analysis, an analysis can be specified in a way that it fits the data by using GLMM (Lo & Andrews, 2015). The GLMM approach allows correlated data structures, non-normal dependent variables, and non-linear relationships between the dependent and independent variables. GLMM is also not sensitive to unequal group sizes, does not require homogeneity of variances, as the distribution can be defined and the error function specified. Therewith, it can be used for data fulfilling the ANOVA assumptions as well, as ANOVA is simply a special case of GLMM. The GLMM approach was taken with both DVs (accuracy of response, response times), for each in a separate model.

More importantly, GLMM was used as the accuracy of response rates, from a statistical point of view, constitute count data as the percentages are calculated on basis of a number of correct responses out of a given amount of trials. More precisely, accuracy of response data constitute proportion data due to the fixed upper limit, which follow a binomial distribution (whereas ANOVA is suitable for

Gaussian distributions). Proportion data ranges from 0 to 1 (i.e. 0 to 100% accuracy). Arcsine transformation, which is recommended for proportion data by some (Wagner, 1993) and been called asinine by others (Warton & Hui, 2011), was refrained from, as GLMM is suitable for proportion data.

The model specifications that were identical for both DVs (accuracy of response and response latencies) are as follows. The data structure included 'subject' and 'sex' as subject variables. The repeated statements were 'intensity' and 'emotion' due to their dependency characteristics. The fixed factors specified in the model were the main effects of 'sex', 'emotion', and 'intensity', as well as the interactions between all factors ('sex*emotion', 'sex*intensity', sex*emotion*intensity). The random intercept included was 'subject' to account for the variability of the intercepts between the subjects. The covariance structure specified was 'diagonal'. Satterthwaite approximation for the degrees of freedom was applied due to the unbalanced design (n(males) = 41, n(females) = 51) and a robust estimation of the fixed effects and coefficients was applied to account for potential model violations. Simple contrasts were retrieved for the comparisons of males and females on the variables included. Pairwise contrasts were retrieved for comparisons of the intensity levels to each other, the emotions to each other, and the intensities within the emotions. Sequential Bonferroni-corrections for the contrasts to correct for multiple comparisons were applied. The results were hence compared to a p-value of .05 after sequential Bonferroni-correction. For DV 1 (accuracy of response), a binomial data structure and logit error function were specified, as appropriate for proportion data (Bolker et al., 2009). When using binomial distribution the resulting output presents the data ranging from 0 to 1. Gamma distribution in combination with log error function was specified for DV 2 (response latencies).

Even though GLMM is suitable for a multitude of data distributions, it can still be subject to outliers. Consequently, the DVs were inspected for extreme values using boxplots. Precisely, the accuracies and response latencies of the emotion categories at each intensity level (27 variables) were inspected for males and females separately. Data transformations were inapplicable, as explained above. For this particular case, Field (2009) suggested the identified extreme values should be replaced by less extreme values. Consequently, the identified extreme values were each changed to the lowest score on the respective variable within their group of sex. This did not change the rank of those cases, but made them less extreme. No accuracy of response data was eliminated.

The extreme values found for the response latencies were corrected the same way as for the accuracy of response data. That is, no data was eliminated, the values were adapted making them less extreme. All extreme values were found at the upper end. The inclusion of correct trials only has led to 101 missing cases (out of 2484 cases). The GLMM was hence conducted with 2383 cases. The GLMM was conducted twice, once with the original values, and once with the corrected values. The binned scatterplots obtained, which should show a diagonal line, showed a visually obvious improvement with the corrected values. With the original values the distribution was clearly non-linear after fitting, but near linear with the corrected values. (Note, the data does not have to be linear for the analysis, but the distribution and link function specification should make the data linear when plotted against predicted values). The results reported below in the results section are from the GLMM with the amended values (the results based on the uncorrected values, which showed the same pattern of findings, are presented in Appendix B for accuracy of response and Appendix C for response times).

Results

Affective state check.

A rating of 3 for the valence item and an arousal rating between 1 and 2 reflected a neutral affective state. Shapiro-Wilk tests showed that the pre- and post-clip ratings for valence and arousal were significantly different from a normal distribution (W’s(92) < .83, p’s < .001). No extreme values were identified using boxplots. To test whether the ratings changed from before to after the neutral clip, Wilcoxon matched pairs signed-rank tests were performed, since all variables were significantly different from a normal distribution. The ratings on valence changed significantly from before (*Mdn* = 4.00, M = 3.76, SD = 0.75) to after (*Mdn* =3.00, M = 3.41, SD = 0.71) the neutral clip, z = -4.33, exact p < .001. The ratings on arousal showed a trend of change from before (*Mdn* = 2.00, *M* = 1.93, *SD* = 0.82) to after (*Mdn* = 2.00, *M* = 1.78, *SD* = 0.81) the neutral clip, $z = -1.99$, $p = .047$, exact $p = .059$. The post-ratings indicate that afterwards the neutral clip the affective state of participants was ‘neutral’.

Males and females were compared regarding their valence and arousal level after watching the neutral clip to test if both groups were in a comparable affective state. The female ($W(51) = .79$) and male ($W(41) = .86$) post-clip valence variable were significantly different from a normal distribution as indicated by Shapiro-Wilk tests (p’s < .001). The female ($W(51) = .80$) and male ($W(41) = .79$) post-clip arousal variable were significantly different from a normal distribution as indicated by Shapiro-Wilk tests (p’s < .001). No extreme values were identified by inspecting of the boxplots. Homogeneity of variances was given for both variables (F’s(1,90) < 2.33, p’s > .130). Based on those results and since the variables were ordinal scaled, Mann-Whitney *U* tests were conducted for the group comparisons. Results showed that males (*Mdn* = 2.00, *SD* = .89) did not differ significantly from females (*Mdn* = 2.00, *SD* = .74) on their post-clip arousal ratings ($U = 1012.50$, $z = -.281$, exact $p = .772$). Males (*Mdn* = 3.00, *SD* = .81) also did not differ from females (*Mdn* = 3.00, *SD* = .64) on their post-clip valence ratings ($U = 1025.00$, $z = -.177$, $p = .865$). Males and females underwent the facial emotion recognition task in comparable affective states.

The 10 emotion categories.

The results from the Wilcoxon signed-rank tests showed that with a Bonferroni-corrected p-value of .005 all 10 variables were recognised significantly above the 10% chance level of responding, all p’s < .001: anger *Mdn* = .77, range (*ran*) = .86 ($z = 8.34$), disgust *Mdn* = .69, *ran* = .97 ($z = 8.32$), fear *Mdn* = .67, *ran* = .81 ($z = 8.33$), sadness *Mdn* = .81, *ran* = .78 ($z = 8.34$), surprise *Mdn* = .94, *ran* = .56 ($z = 8.37$), happiness *Mdn* = .89, *ran* = .58 ($z = 8.34$), neutral *Mdn* = .92, *ran* = .81 ($z = 8.35$), contempt *Mdn* = .31, *ran* = .97 ($z = 6.23$), pride *Mdn* = .47, *ran* = .94 ($z = 7.19$) and embarrassment *Mdn* = .67, *ran* = .86 ($z = 8.33$); see Table 4 for means and standard deviations in decimal fractions.

Table 4

Accuracy of Responses for the 10 Emotion Categories

Emotion (*N* = 92)	**Means**	**Standard Deviations**
Anger	.74	.18
Disgust	.65	.23
Fear	.62	.20
Sadness	.79	.15
Surprise	.92	.10
Happiness	.84	.14
Neutral	.89	.13
Pride	.42	.27
Embarrassment	.65	.17
Contempt	.34	.24

Note. Means and standard deviations are expressed in decimal fractions.

The nine emotions at each intensity level.

The results from the Wilcoxon signed-rank tests showed that the median accuracy of responses for each of the 27 categories were recognised significantly above chance level of responding (10%) with a Bonferroni-corrected *p*-value of .002 (all *p*'s < .001): anger low intensity *Mdn* = .67, *ran* = 1.00 (*z* = 8.29), anger intermediate intensity *Mdn* = .83, *ran* = 1.00 (*z* = 8.36), anger high intensity *Mdn* = .92, *ran* = .92 (*z* = 8.40), disgust low intensity *Mdn* = .67, *ran* = .92 (*z* = 8.32), disgust intermediate intensity *Mdn* = .67, *ran* = 1.00 (*z* = 8.32), disgust high intensity *Mdn* = .79, *ran* = 1.00 (*z* = 8.33), fear low intensity *Mdn* = .50, *ran* = 1.00 (*z* = 8.24), fear intermediate intensity *Mdn* = .67, *ran* = .92 (*z* = 8.34), fear high intensity *Mdn* = .75, *ran* = .92 (*z* = 8.34), sadness low intensity *Mdn* = .75, *ran* = .75 (*z* = 8.36), sadness intermediate intensity *Mdn* = .83, *ran* = .75 (*z* = 8.40), sadness high intensity *Mdn* = .83, *ran* = .92 (*z* = 8.37), surprise low intensity *Mdn* = .92, *ran* = .67 (*z* = 8.44), surprise intermediate intensity *Mdn* = 1.00, *ran* = .83 (*z* = 8.49), surprise high intensity *Mdn* = 1.00, *ran* = .42 (*z* = 8.64), happiness low intensity *Mdn* = .75, *ran* = 1.00 (*z* = 8.31), happiness intermediate intensity *Mdn* = .92, *ran* = .75 (*z* = 8.45), happiness high intensity *Mdn* = 1.00, *ran* = .50 (*z* =

8.67), contempt low intensity *Mdn* = .17, *ran* = .92 (z = 5.00), contempt intermediate intensity *Mdn* = .33, *ran* = 1.00 (z = 5.99), contempt high intensity *Mdn* = .42, *ran* = 1.00 (z = 6.67), pride low intensity *Mdn* = .25, *ran* = .83 (z = 6.57), embarrassment low intensity *Mdn* = .42, *ran* = .83 (z = 8.28), embarrassment intermediate intensity *Mdn* = .67, *ran* = .92 (z = 8.32), embarrassment high intensity *Mdn* = .92, *ran* = .92 (z = 8.38), pride intermediate intensity *Mdn* = .50, *ran* = 1.00 (z = 7.15), and pride high intensity *Mdn* = .58, *ran* = 1.00 (z = 7.30). The means and standard deviations for the accuracy of responses of the emotion categories at each intensity level of expression are given in Table 5.

Table 5

Accuracy of Responses for the Emotion Categories by Intensity

Emotion	**Means (Standard Deviations)**		
(*N* = 92)	**low**	**intermediate**	**high**
Anger	.60 (.23)	.79 (.21)	.85 (.18)
Sadness	.72 (.19)	.82 (.14)	.84 (.17)
Disgust	.58 (.25)	.66 (.25)	.71 (.24)
Fear	.51 (.24)	.63 (.24)	.71 (.23)
Happiness	.68 (.27)	.90 (.15)	.96 (.08)
Surprise	.90 (.13)	.92 (.12)	.95 (.09)
Contempt	.27 (.18)	.37 (.32)	.41 (.32)
Embarrassment	.46 (.19)	.63 (.21)	.85 (.19)
Pride	.30 (.22)	.45 (.31)	.52 (.34)

Note. The overall mean accuracy for all videos together was .69 (*SD* = .09). Means and standard deviations are expressed as decimal fractions.

Individual video analysis.

The results of the Binomial test showed that of the 360 videos, for 350 videos the recognition rates were significantly greater than the 10% chance level of responding with a Bonferroni-corrected p-value of .000139 (proportions correct > .30, exact p's < . 000139; 1-tailed). Instead of listing individual results of 360 videos, only the results of the 10 videos that were not significantly different from chance level of responding are listed here: F03 disgust low intensity (proportion correct = .20, exact p = .019845), F01 contempt low intensity (proportion correct = .20, exact p = .000248), F05 contempt low intensity (proportion correct = .10, exact p = .206932), M02 contempt low intensity (proportion correct = .20, exact p = .019845), M08 contempt low intensity (proportion correct = .10, exact p = .127664), M04 pride low intensity (proportion correct = .10, exact p = .312624), F04 pride low intensity (proportion correct = .20, exact p = .001716), M06 pride low intensity (proportion correct = .10, exact p = .092176), M12 pride low intensity (proportion correct = .20, p = .004131), and M12 embarrassment low intensity (proportion correct = .10, exact p = .206932). The remaining 350 videos were recognised significantly above chance level of responding (proportions correct > .20, p's < .000139). (See Appendix D for full table of accuracy of response per video).

GLMM with accuracy of response as the DV.

The results from the GLMM for the accuracy rates showed a significant main effect of expression *intensity* (F(2,748) = 435.50, p < .001). The pairwise contrasts showed that the accuracy of responses for the low intensity level (M = .57, SE = .01) were significantly lower than for the intermediate intensity level (M = .73, SE = .01; t(834) = -20.08, p < .001, after sequential Bonferroni-correction), which in turn were significantly lower than for the high intensity level (M = .81, SE = .01; t(679) = -12.09, p < .001, after sequential Bonferroni-correction); see Figure 9.

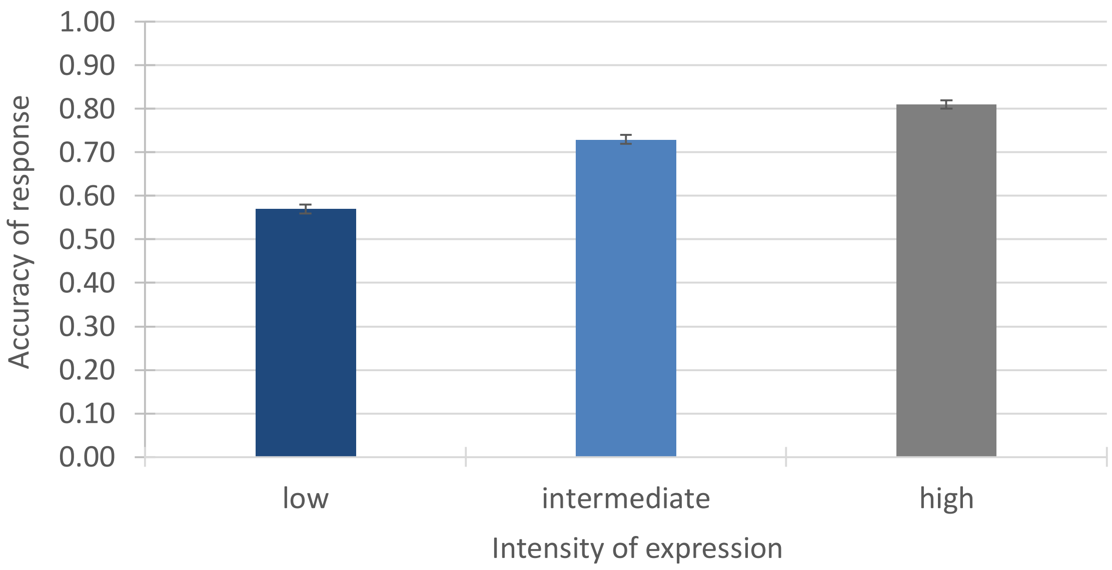


Figure 9. Accuracy of response for the three intensity levels of facial expression of the ADFES-BIV. Accuracy of responses are expressed as decimal fractions. Chance level of responding = .10. Error bars represent standard errors of the means.

The main effect of *emotion* was significant (F(8,85) = 60.63, p < .001). The rank order of recognition from highest to lowest was: surprise (M = .93, SE = .01), happiness (M = .89, SE = .01), sadness (M = .80, SE = .01), anger (M = .77, SE = .01), embarrassment (M = .67, SE = .02), disgust (M = .65, SE = .02), fear (M = .62, SE = .02), pride (M = .42, SE = .03), and contempt (M = .33, SE = .03); see Figure 10. (Note, these means deviate minimally from the reports of the analysis above based on the original data, since here extreme value adjustment was carried out). Pairwise contrasts showed that most emotions were significantly different from each other in accuracy of response (t's(31-158) = 2.82-18.97, p's ≤ .032, after sequential Bonferroni-correction). The emotions that did not differ significantly from each other in accuracy of response were anger and sadness (t(61) = -1.60, p = .370, after sequential Bonferroni-correction), disgust and embarrassment (t(69) = -0.60, p = .631, after sequential Bonferroni-correction), disgust and fear (t(67) = 1.02, p = .631, after sequential Bonferroni-correction), and embarrassment and fear (t(72) = 1.71, p = .370, after sequential Bonferroni-correction). A trend was found for the difference in accuracy of response between contempt and pride (t(90) = -2.49, p = .073, after sequential Bonferroni-correction).

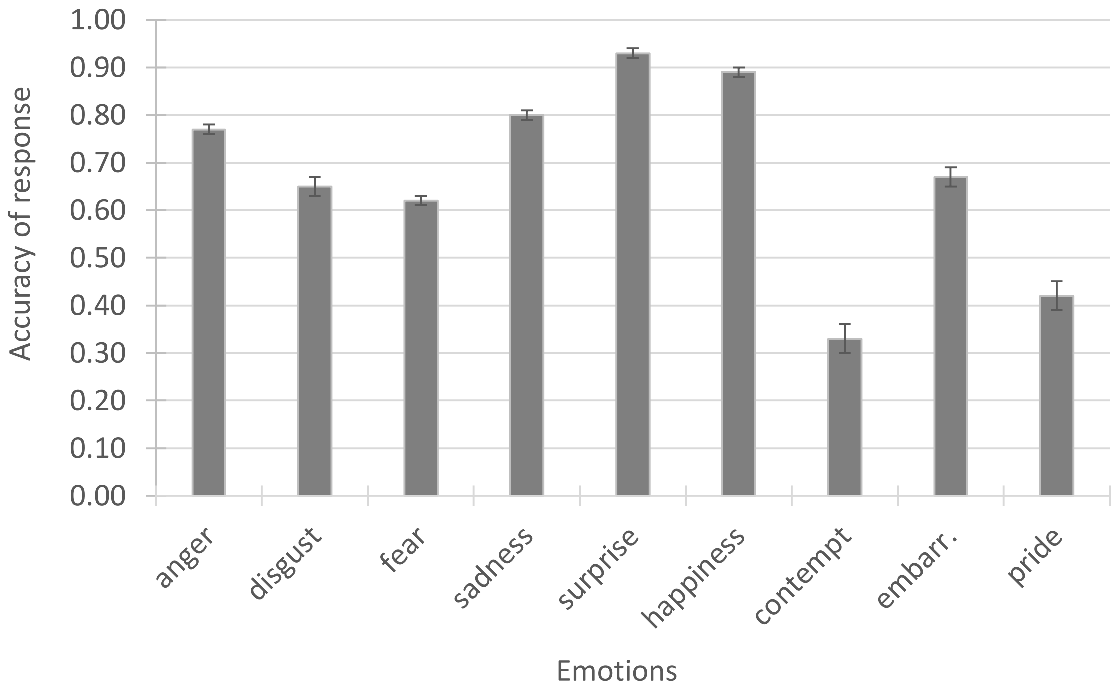


Figure 10. Accuracy of response for the emotion categories of the ADFES-BIV. Accuracy of responses are expressed as decimal fractions. Chance level of responding = .10. Embarr. = embarrassment. Error bars represent standard errors of the means.

The interaction of *emotion*intensity* was significant ($F(16,557) = 22.34$, $p < .001$). The rank order of emotion recognition varied for the intensities; see Figure 11. Pairwise contrasts showed that only the accuracy of responses for the intermediate compared to high intensity expression of sadness were not significantly different from each other ($t(171) = -0.71$, $p = .484$, after sequential Bonferroni-correction). All other expression intensity levels within the emotion categories were significantly different from each other (t(124-1181) = 1.97-22.89, p's ≤ .05, after sequential Bonferroni-correction), with increasing accuracy of response with increasing expression intensity level. Means and *SE*s are given in Table 6.

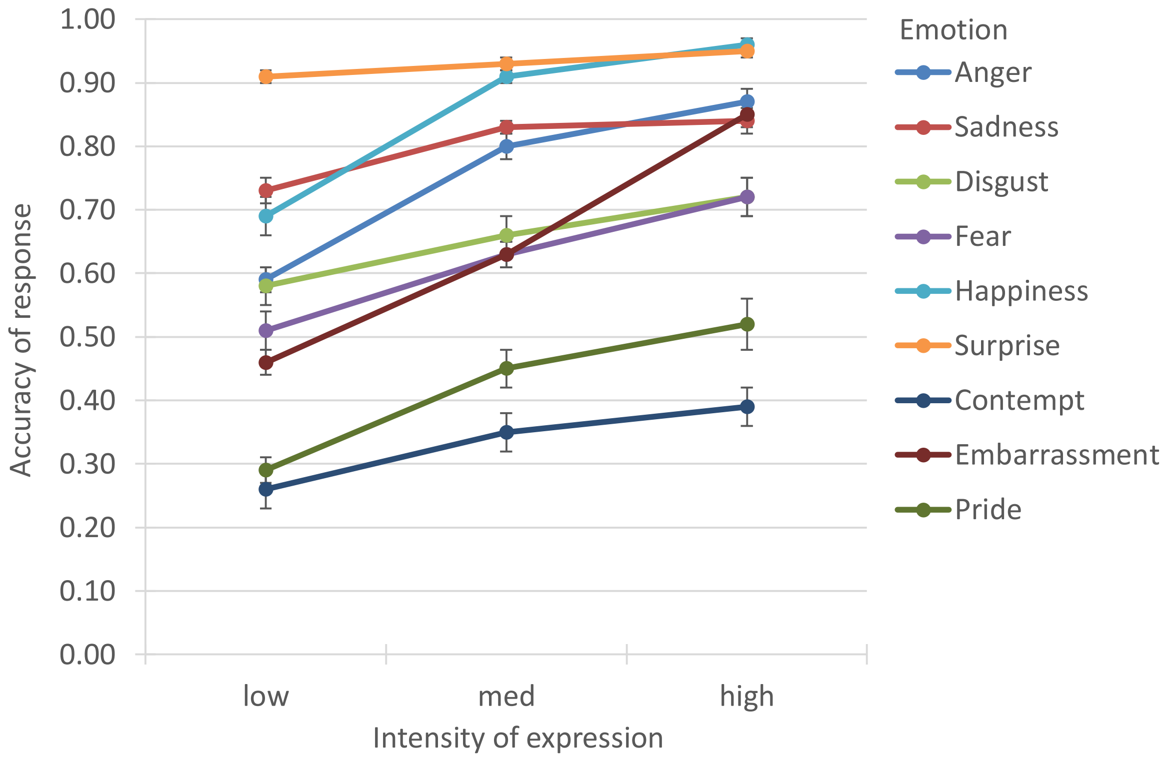


Figure 11. Accuracy of response for the emotion categories of the ADFES-BIV at each intensity level. Accuracy of responses are expressed as decimal fractions. Chance level of responding = .10. Error bars represent standard errors of the means.

Table 6

Accuracy of Responses for the Emotion Categories by Intensity

Emotion	**low**		**intermediate**		**high**	
(*N* = 92)	***M***	***SE***	***M***	***SE***	***M***	***SE***
Anger	.59	.02	.80	.02	.87	.02
Sadness	.73	.02	.83	.01	.84	.02
Disgust	.58	.03	.66	.03	.72	.03
Fear	.51	.03	.63	.02	.72	.03
Happiness	.69	.03	.91	.01	.96	.01
Surprise	.91	.01	.93	.01	.95	.01
Contempt	.26	.03	.35	.03	.39	.03
Embarrassment	.46	.02	.63	.02	.85	.02
Pride	.29	.02	.45	.03	.52	.04

Note. Means (*M*) and standard errors of the means (*SE*) are expressed in decimal fractions.

The main effect of *sex* was significant ($F(1,94) = 15.07$, $p < .001$) with females (M = .75, SE = .01) outperforming males (M = .67, SE = .02) based on accuracy of response. The interaction of *sex*intensity* was also significant ($F(2,748) = 3.91$, $p = .020$). The simple contrast showed that females (M = .61, SE = .02) outperformed males (M = .54, SE = .02) at recognising low intensity expressions ($t(122) = 2.61$, $p = .010$, after sequential Bonferroni-correction), females (M = .77, SE = .01) outperformed males (M = .68, SE = .02) at recognising intermediate intensity expressions ($t(100) = 4.18$, $p < .001$, after sequential Bonferroni-correction), and females (M = .84, SE = .01) outperformed males (M = .77, SE = .02) at recognising high intensity expressions ($t(164) = 3.75$, $p < .001$, after sequential Bonferroni-correction) with the difference being greatest at intermediate intensity (9% vs 7% at low and 7% at high intensity); see Figure 12.

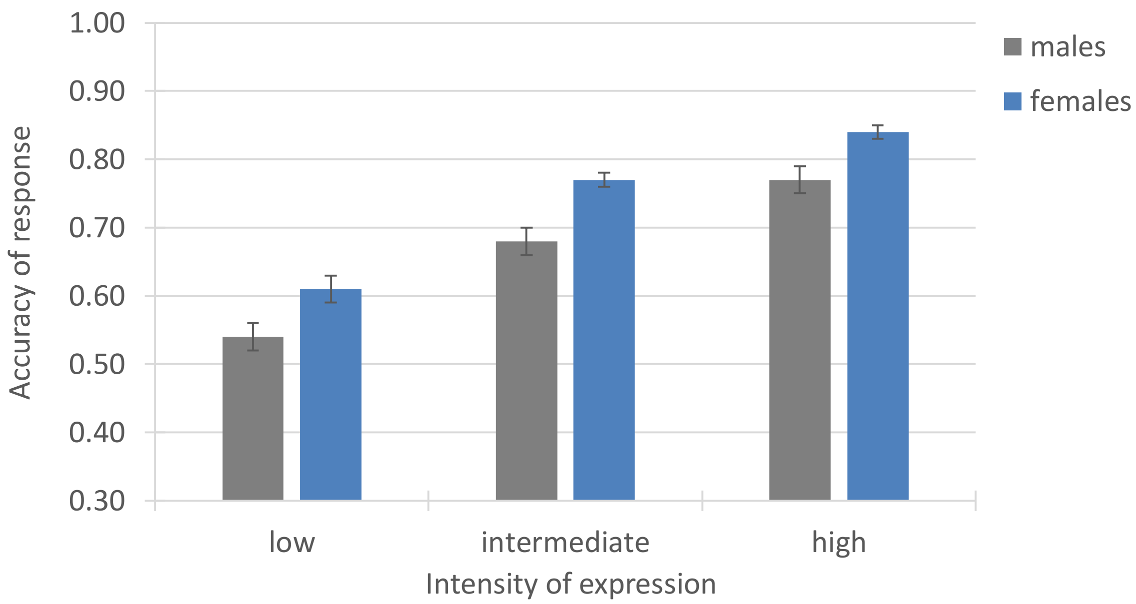


Figure 12. Accuracy of response of the ADFES-BIV intensity levels for males and females. Accuracy of responses are expressed as decimal fractions. Error bars represent standard errors of the means.

The interaction of *sex*emotion* was not significant ($F(8,85) = 1.59$, $p = .140$). The interaction of *sex*emotion*intensity* was significant ($F(16,557) = 1.85$, $p = .023$). That is, the accuracy of responses for males and females were differently influenced by intensity of expression across the varying emotions; see Figure 13 (means and their standard errors are given in Table 7). At low intensity, females outperformed males at recognition of anger

($t(63) = 2.01$, $p = .049$, after sequential Bonferroni-correction), disgust ($t(68) = 2.57$, $p = .012$, after sequential Bonferroni-correction), embarrassment ($t(98) = 2.00$, $p = .048$, after sequential Bonferroni-correction), and sadness ($t(101) = 2.81$, $p = .006$, after sequential Bonferroni-correction). At intermediate intensity, females outperformed males at recognition of anger ($t(33) = 2.96$, $p = .006$, after sequential Bonferroni-correction), contempt ($t(35) = 2.43$, $p = .020$, after sequential Bonferroni-correction), disgust ($t(71) = 2.40$, $p = .019$, after sequential Bonferroni-correction), fear ($t(78) = 1.99$, $p = .050$, after sequential Bonferroni-correction), happiness ($t(112) = 2.85$, $p = .005$, after sequential Bonferroni-correction), sadness ($t(87) = 3.69$, $p < .001$, after sequential Bonferroni-correction), and embarrassment ($t(79) = 2.00$, $p = .049$, after sequential Bonferroni-correction). At high intensity, females only scored significantly higher than males for anger recognition ($t(34) = 3.37$, $p = .002$, after sequential Bonferroni-correction) and recognition of contempt ($t(38) = 2.83$, $p = .007$, after sequential Bonferroni-correction). Trends for a female advantage over males were found for disgust recognition ($t(67) = 1.88$, $p = .064$, after sequential Bonferroni-correction) and fear recognition ($t(73) = 1.86$, $p = .067$, after sequential Bonferroni-correction).

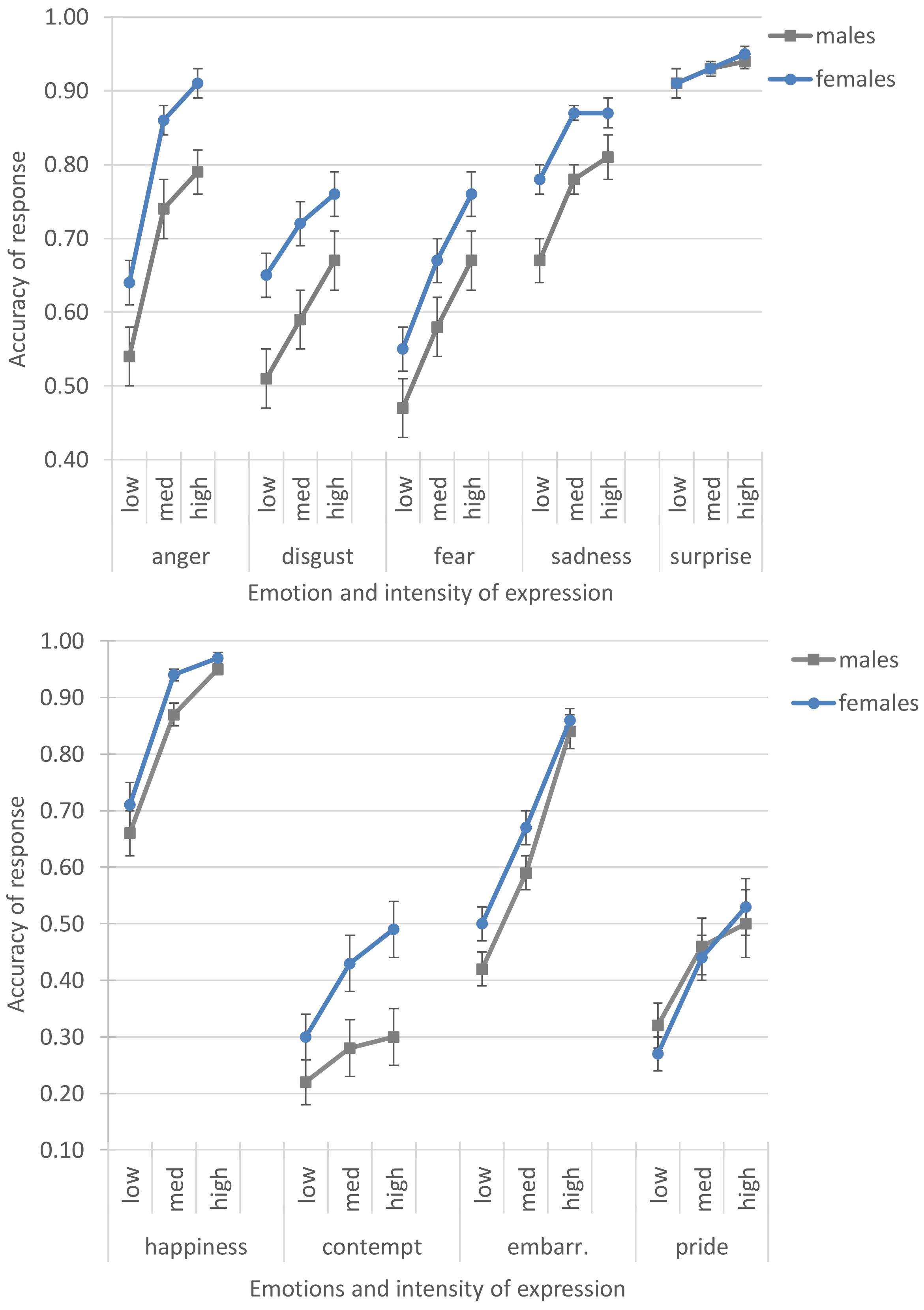


Figure 13. Accuracy of responses of the ADFES-BIV emotion categories by intensity of expression for males and females. Accuracy of responses are expressed as decimal fractions. Embarr. = embarrassment. Error bars represent standard errors of the means.

Table 7

Accuracy of Responses for the Emotion Categories by Intensity for Males and Females

	Males (*n* = 41)						**Females (*n* = 51)**					
	low		**intermed.**		**high**		**low**		**intermed.**		**high**	
Emotion	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
Anger	.54	.04	.74	.04	.79	.03	.64	.03	.86	.02	.91	.01
Disgust	.51	.04	.59	.04	.67	.04	.65	.03	.72	.03	.76	.03
Fear	.47	.04	.58	.04	.67	.04	.55	.03	.67	.03	.76	.03
Sadness	.67	.03	.78	.02	.80	.03	.78	.02	.87	.01	.87	.02
Surprise	.91	.02	.93	.01	.94	.01	.91	.02	.93	.01	.95	.01
Happiness	.66	.04	.87	.02	.95	.01	.71	.04	.94	.01	.97	.01
Contempt	.22	.04	.28	.05	.30	.05	.30	.04	.43	.05	.49	.05
Embarr.	.42	.03	.59	.03	.84	.03	.50	.03	.67	.03	.86	.02
Pride	.32	.04	.46	.05	.50	.06	.27	.03	.44	.04	.53	.05

Note. Means (*M*) and standard errors of the means (*SE*) are expressed in decimal fractions. Embarr. = embarrassment. Intermed. = intermediate.

GLMM with response latencies as the DV.

The main effect of expression *intensity* was significant ($F(2,271) = 98.08$, $p < .001$). Paired contrasts showed that participants were 125ms slower in recognising expressions at low intensity ($M = 1096$, $SE = 32$) than at intermediate intensity ($M = 971$, $SE = 29$, $t(487) = 7.44$, $p < .001$, after sequential Bonferroni-correction) and 102ms slower at recognising the latter than the high intensity level ($M = 869$, $SE = 23$, $t(487) = 6.45$, $p < .001$, after sequential Bonferroni-correction), see Figure 14.

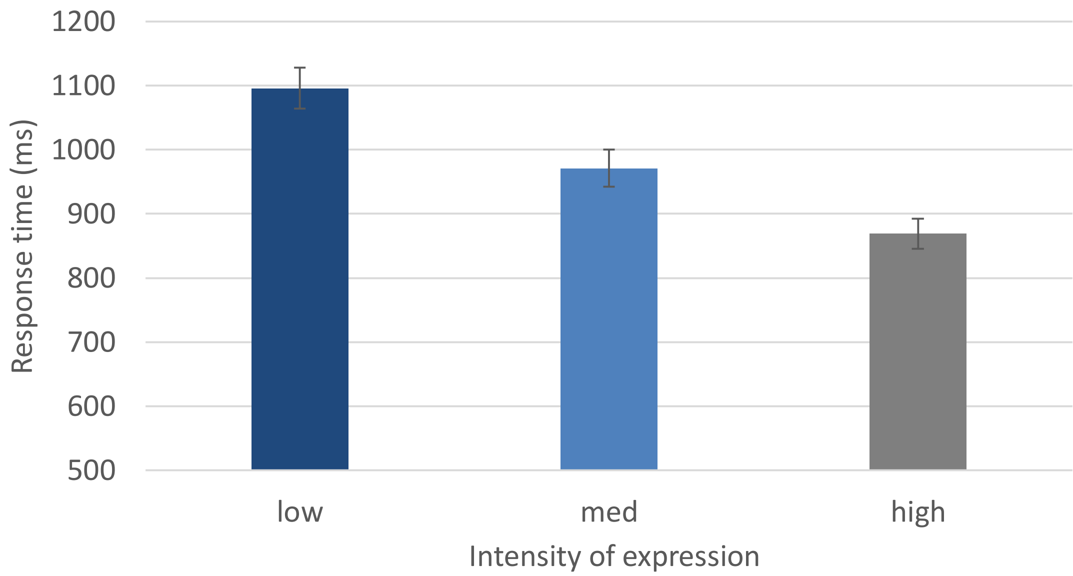


Figure 14. Response times (ms) for the three intensity levels of facial expression of the ADFES-BIV. Error bars represent standard errors of the means.

The main effect of *emotion* was significant ($F(8,148) = 26.99$, $p < .001$). The rank order of recognition, with a difference of 72ms between the emotion category fastest recognised and slowest recognised, was as follows: happiness ($M = 803$, $SE = 25$), surprise ($M = 823$, $SE = 25$), pride ($M = 867$, $SE = 35$), sadness ($M = 891$, $SE = 33$), disgust ($M = 956$, $SE = 39$), anger ($M = 964$, $SE = 35$), embarrassment ($M = 1003$, $SE = 39$), fear ($M = 1123$, $SE = 40$), contempt ($M = 1488$, $SE = 76$), see Figure 15. Paired contrasts showed that *happiness* was recognised significantly faster than anger ($t(101) = -6.09$, $p < .001$, after sequential Bonferroni-correction), contempt ($t(71) = -9.54$, $p < .001$, after sequential Bonferroni-correction), disgust ($t(57) = -4.53$, $p = .001$, after sequential Bonferroni-correction), embarrassment ($t(75) = -5.14$, $p < .001$, after sequential Bonferroni-correction), fear ($t(102) = -8.46$, $p < .001$, after sequential Bonferroni-correction), and sadness (t's$(115) = -3.40$, $p = .014$, after sequential Bonferroni-correction), but not surprise ($t(126) = -0.89$, $p = 1.00$, after sequential Bonferroni-correction). *Surprise* was recognised significantly faster than anger ($t(80) = -4.67$, $p < .001$, after sequential Bonferroni-correction), contempt ($t(73) = -9.54$, $p < .001$, after sequential Bonferroni-correction), disgust ($t(54) = -3.50$, $p = .004$, after sequential Bonferroni-correction), embarrassment ($t(117) = -5.34$, $p < .001$, after sequential Bonferroni-correction), and fear ($t(116) = -7.94$, $p < .001$, after sequential Bonferroni-correction), but not sadness ($t(88) = -2.54$, $p = .153$, after sequential Bonferroni-correction) and pride ($t(103) = -1.28$, $p = 1.00$, after sequential Bonferroni-correction). *Pride* was significantly faster recognised than contempt ($t(80) = -8.15$, $p < .001$, after sequential Bonferroni-correction), embarrassment ($t(194) = -3.52$, $p = .009$, after sequential Bonferroni-correction), and fear ($t(132) = -5.53$, $p < .001$, after sequential Bonferroni-correction), but not sadness ($t(106) = -0.65$, $p = 1.00$, after sequential Bonferroni-correction), disgust ($t(75) = -2.08$, $p = .372$, after sequential Bonferroni-correction), and anger ($t(102) = -2.45$, $p = .175$, after sequential Bonferroni-correction). *Sadness* was recognised significantly faster than contempt ($t(83) = -8.55$, $p < .001$, after sequential Bonferroni-correction) and fear ($t(96) = -5.61$, $p < .001$, after sequential Bonferroni-correction), but not disgust ($t(102) = -1.81$, $p = .592$, after sequential Bonferroni-correction) and anger ($t(141) = -2.26$, $p = .256$, after sequential Bonferroni-correction). A trend was found for the difference between sadness and embarrassment ($t(99) = -2.92$, $p = .056$, after sequential Bonferroni-correction). *Disgust* was significantly faster recognised than fear ($t(144) = -4.55$, $p < .001$, after sequential Bonferroni-correction) and contempt ($t(101) = -8.15$, $p < .001$, after sequential Bonferroni-correction), but not anger ($t(101) = 6.09$, $p < .001$, after sequential Bonferroni-correction) and embarrassment ($t(79) = -0.21$, $p = 1.00$, after sequential Bonferroni-correction). *Anger* was significantly faster recognised than fear ($t(89) = -3.98$, $p = .003$, after sequential Bonferroni-correction) and contempt ($t(82) = -7.46$, $p < .001$, after sequential Bonferroni-correction), but not embarrassment ($t(78) = -0.93$, $p = 1.00$, after sequential Bonferroni-correction). *Embarrassment* was significantly faster recognised than contempt ($t(95) = -6.81$, $p < .001$, after sequential Bonferroni-correction) and fear ($t(182) = -3.08$, $p = .033$, after sequential Bonferroni-correction). *Fear* was significantly faster recognised than contempt ($t(90) = -5.30$, $p < .001$, after sequential Bonferroni-correction).

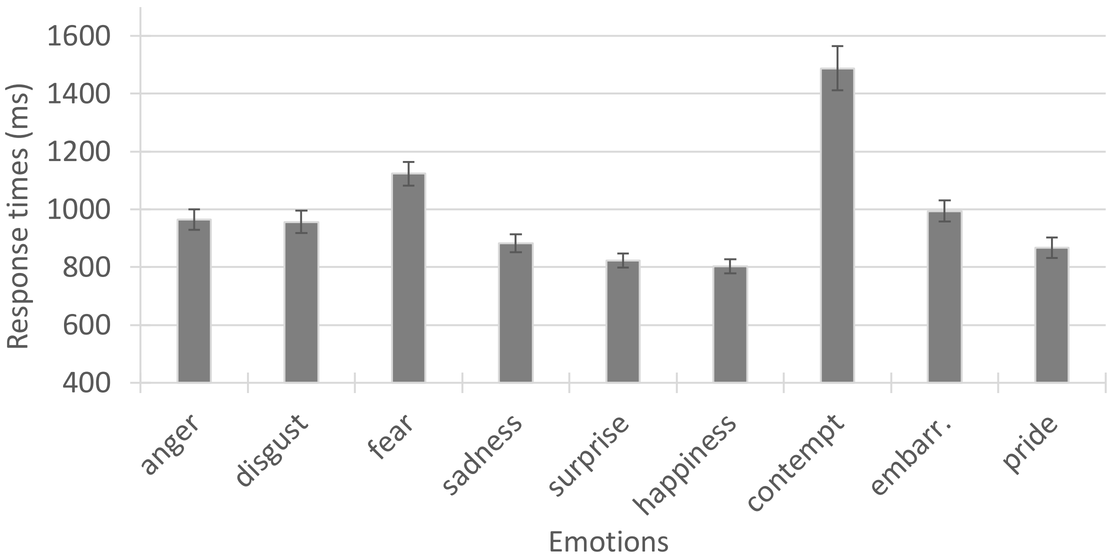


Figure 15. Response times (ms) for the emotion categories of the ADFES-BIV. Embarr. = embarrassment. Error bars present the standard errors of the means.

The interaction *emotion*intensity* was significant ($F(16,349) = 9.82$, $p < .001$), see Figure 16 (means and their standard errors are presented in Table 8). Low intensity expressions of *happiness* were recognised significantly slower than intermediate intensity expressions of happiness ($t(106) = 4.48$, $p < .001$, after sequential Bonferroni-correction) and the latter were recognised significantly slower than high intensity happiness expressions ($t(117) = 7.02$, $p < .001$, after sequential Bonferroni-correction). Recognition of *surprise* at low intensity occurred significantly slower than recognition of surprise at intermediate intensity ($t(125) = 3.13$, $p = .002$, after sequential Bonferroni-correction) and the latter was slower recognised than surprise at high expression intensity ($t(57) = 4.03$, $p < .001$, after sequential Bonferroni-correction). The response times for the low intensity expressions of *anger* were significantly lower than for the intermediate intensity expressions of anger ($t(100) = 6.31$, $p < .001$, after sequential Bonferroni-correction) and a trend was found for the difference in response times for the latter and high intensity anger expressions ($t(178) = 1.85$, $p = .066$, after sequential Bonferroni-correction). Recognition of *fear* was significantly faster at high than intermediate intensity ($t(70) = -3.92$, $p < .001$, after sequential Bonferroni-correction), but the response time was not significantly different between the low and intermediate intensity of fear ($t(74) = 1.40$, $p = .167$, after sequential Bonferroni-correction). For *pride*, the low expression intensity was significantly slower recognised than the intermediate intensity ($t(170) = 5.08$, $p < .001$, after sequential Bonferroni-correction), but the response times for recognition of the intermediate intensity was not significantly different from the high intensity ($t(229) = -1.53$, $p = .127$, after sequential Bonferroni-correction). For *sadness*, the response time for recognising intermediate intensity expressions was not significantly different from the high intensity ($t(135) = 0.73$, $p = .470$, after sequential Bonferroni-correction), but the response time for recognising low intensity expressions was significantly different from the intermediate intensity ($t(215) = 2.44$, $p = .031$, after sequential Bonferroni-correction). For *disgust*, the decrease in response time from intermediate to high intensity was significantly different ($t(151) = 4.04$, $p < .001$, after sequential Bonferroni-correction) and for the difference in response times from low to intermediate intensity a trend was found ($t(132) = 1.77$, $p = .079$, after sequential Bonferroni-correction). Recognition of *embarrassment* occurred significantly faster at intermediate intensity than at low intensity ($t(161) = -2.55$, $p = .012$, after sequential Bonferroni-correction) and significantly at high intensity than at intermediate intensity ($t(487) = -4.68$, $p < .001$, after sequential Bonferroni-correction). The differences in response time for recognition of *contempt* at low intensity compared to intermediate intensity was not significantly different ($t(145) = -1.87$, $p = .190$, after sequential Bonferroni-correction), neither was the difference in response time to contempt at intermediate intensity compared to high intensity ($t(212) = 0.92$, $p = .654$, after sequential Bonferroni-correction).

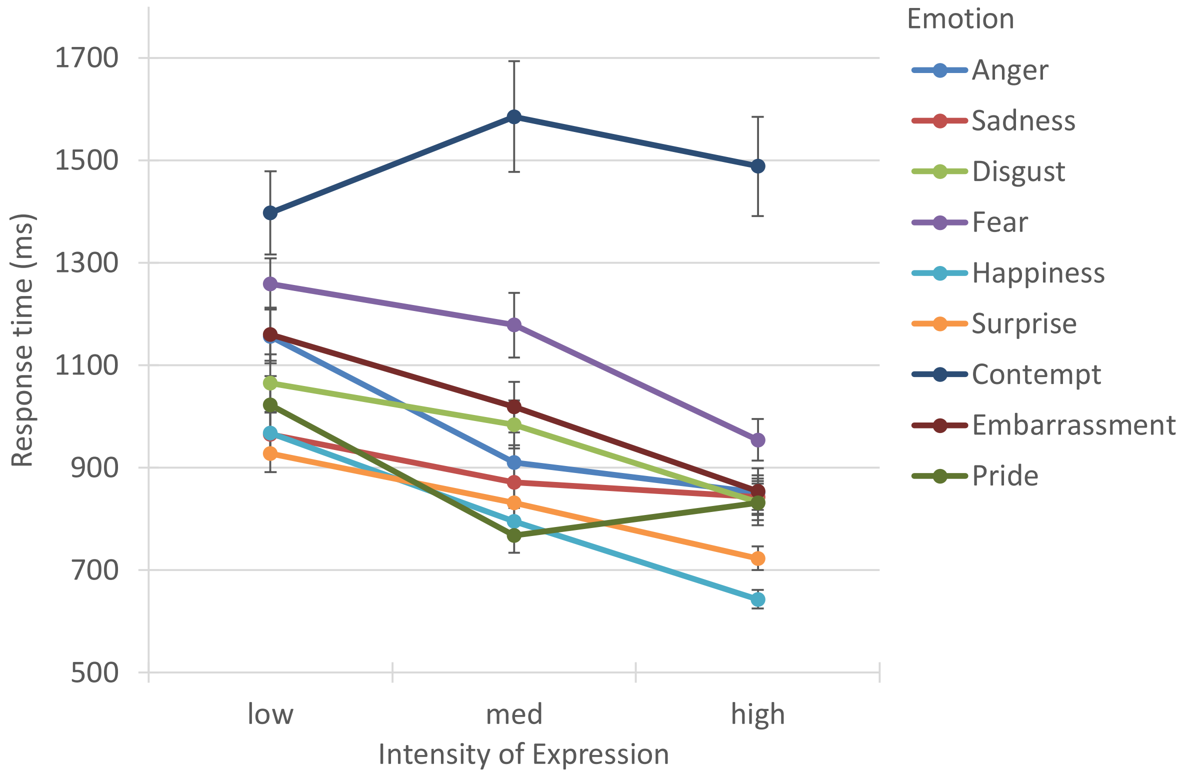


Figure 16. Response times (ms) for the emotion categories of the ADFES-BIV by intensity. Error bars present the standard errors of the means.

Table 8

Response Times for the ADFES-BIV Emotion Categories by Intensity

Emotion	**low**		**intermediate**		**high**	
(*N* = 92)	***M***	***SE***	***M***	***SE***	***M***	***SE***
Anger	1156	53	910	34	851	34
Sadness	965	43	871	40	843	36
Disgust	1065	56	984	47	833	36
Fear	1259	49	1178	63	954	41
Happiness	967	40	795	26	643	18
Surprise	927	36	831	32	723	23
Contempt	1397	81	1585	108	1488	97
Embarrassment	1160	52	1018	49	854	44
Pride	1022	57	768	34	831	43

Note. Means (*M*) and standard errors of the means (*SE*) are expressed in ms.

The main effect of *sex* was significant ($F(1,95) = 7.64$, $p = .007$) with females ($M = 904$, $SE = 33$) responding 145ms faster than males ($M = 1049$, $SE = 42$), see Figure 17. The interactions of *sex*emotion* ($F(8, 148) = 1.23$, $p = .283$), *sex*intensity* ($F(2,271) = 0.48$, $p = .622$), and *sex*emotion*intensity* ($F(16,349) = 1.30$, $p = .195$) were not significant.

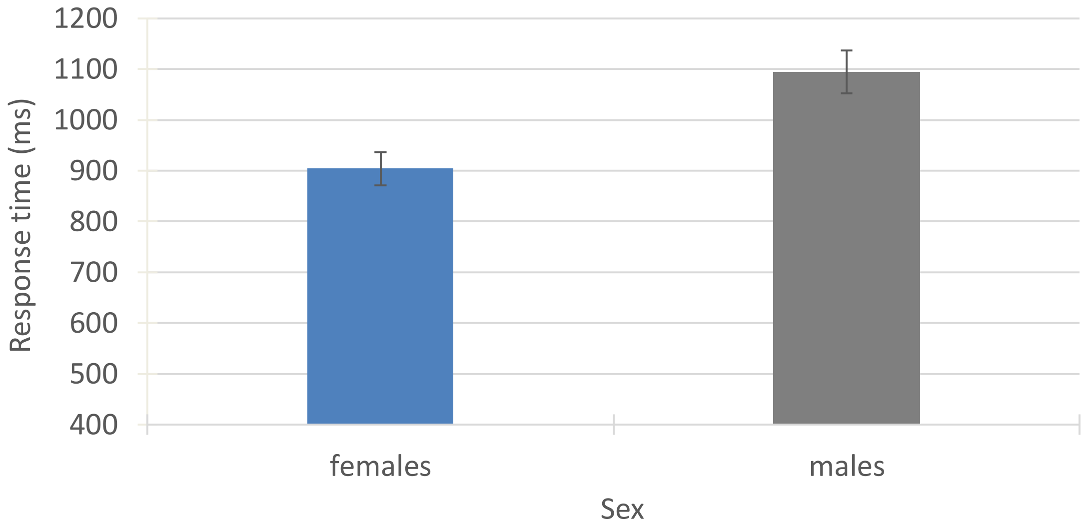


Figure 17. Overall response times (ms) for males and females. Error bars present the standard errors of the means.

Discussion

The main purpose of this study was to validate the ADFES-BIV based on the recognisability of all its categories, which included the intensity levels, emotion categories, emotion categories at the varying intensity levels, and the individual videos. In line with the hypotheses, the ADFES-BIV was shown to be validated for all its categories with recognition rates significantly above the chance level of responding. On the level of individual videos, 97% of the total 360 videos were recognised significantly above the chance level of responding, which aligns with the expectation that at least 96% of the videos would reach accuracy rates above chance. To further validate the ADFES-BIV on basis of its ability to differentiate between individuals' facial emotion recognition ability, sex differences in accuracy of response and response latencies were investigated. In line with the hypothesis, a female advantage compared to males was found regarding accuracy of response and response times, which also demonstrates the videos' high sensitivity to reveal group differences within the general population. Together, the results showed that this newly created video set of different intensities of emotional expressions of basic and complex emotions is a valid set of stimuli for use in research on emotion processing.

Validation.

This study aimed to validate all categories of the ADFES-BIV and the individual videos themselves. In line with the prediction, the recognition rates of the 10 emotion categories, the 27 emotion categories at each of the three intensity levels, and the three intensity levels were found to be significantly greater than the chance level of responding of 10%. The results thereby show that the ADFES-BIV was effectively validated on all its categories. On individual video level, only 10 out of the total 360 videos showed accuracy rates at or below chance level of responding. Those 10 videos all belonged to the low intensity category, with most of them from contempt and pride. Therefore, only 2.8% videos within the whole set were not recognised significantly above chance level of responding, which is in line with the expectation that at least 96% of the videos should be recognised above chance level of responding. It has to be noted that a very conservative *p*-value (.0001) was applied. Had the typical alpha-level of 5% been applied, only 5 videos (= 1.4%) would have failed the significance tests. This finding hence constitutes an improvement to the results from Chapter 4 and indicates a satisfying result on individual video level.

Aim of the current research was also to test the validity of the three created intensity levels by comparing them to each other. Results showed differences between levels of intensity of the expressions, with the lowest mean accuracy of response for the low intensity expressions (57%) compared to the intermediate intensity expressions (73%), which were lower than the high intensity expressions (82%). The linear increase of approximately 10% in accuracy of response between intensity levels is in line with previous research from morphed static stimuli looking at varying intensities where accuracy of response was linearly associated with physical expression intensity (see Hess et al., 1997; Hoffmann et al., 2010; Montagne et al., 2007) and has been shown for the first time from dynamic stimuli, i.e. video recordings. The current study further found that the intensity of facial expressions has an influence on response times, as responses were 126ms slower for recognition at low than intermediate expression intensity and 96ms slower at intermediate than high expression intensity. Fastest responses were given to high intensity expressions and slowest responses to low intensity expressions, in line with the prediction. The lower response times in higher intensities are consistent with the notion that responding occurs more rapidly the easier a stimulus is to recognise (Kestenbaum & Nelson, 1992). This effect of difficulty is further reflected in the negative relationship that is generally found between accuracy of response and response latencies (e.g. Derntl et al., 2009). Also in the current study response latencies decreased with increasing accuracy of response ($r(92) = -.311$, $p = .003$). That is, the longer an individual took to respond, the more likely it was for them to give the wrong answer.

The intensity of the muscle contractions underlying the facial emotional expressions and the amount of facial action units activated constitute a determining factor regarding ease of recognition. That is, the higher the intensity of muscle contraction, the more apparent the facial features that form the emotional facial expressions become, which facilitates recognition. This facilitation is seen both, in accuracy rates and response times, and explains the found differences between the expression intensity levels within the current research. At emotion onset, not all facial action units are activated (e.g. Hess et al., 1997; Jack et al., 2014) and the ones that are activated have not reached their maximum yet, which makes it more difficult to recognise subtle facial emotional expressions. Since subtler displays of facial emotion are frequently encountered in social interactions, it might seem surprising that we are less good at recognising them than expressions of higher intensity. However, that subtler emotional expressions are less well recognised than intense expressions, even though they are a substantial part of our social interactions, can be explained evolutionary. With facial emotional expressions serving the communicative function of signalling threat (among others) (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003), it can be assumed that the level of threat correlates with the intensity of the expression, i.e. the more intense the expression, the more important to recognise the expression. With subtler displays of facial emotion being less relevant for securing survival, good facial emotion recognition skills do not have to extent to subtler levels. A reason why subtle emotional facial expressions are harder to recognise than more intense expressions can also be found in the intention of the encoder. In some situations people mask their feelings based on social norms influenced by culture, context, personality etc. (Ekman, 1971) leading to only subtle facial expressions (intentionally) hard to recognise. In this case, it can actually be beneficial in the interaction of two people for their relationship to not perceive and recognise the emotion. This benefit for the relationship between humans might further explain why we are not equipped to be as effective at recognising low intensity expressions as at recognising high intensity ones. In this context, it would be interesting to examine if highly empathic people are yet better than low empathic people at recognising subtle emotional expressions; a question that could be investigated with the ADFES-BIV.

Regarding the emotion categories, it was hypothesised that the non-negative basic emotions would be better recognised than the negative basic emotions and the complex emotions would be hardest to recognise. Measured on accuracy of response rates, surprise and happiness were found easiest to recognise, fear hardest (from the basic emotions), followed by the complex emotions with contempt the most difficult to recognise. The findings of the current study are hence in line with the hypothesis. Furthermore, the results are in line with the reported order of recognition of the emotion categories by the developers of the ADFES (van der Schalk et al., 2011) and fit in with other published literature where surprise and happiness are usually reported as the emotions easiest to recognise (e.g. Tottenham et al., 2009). A potential reason for the high recognisability of happiness is that the facial expression of happiness is very distinct from all other basic emotions, as it is the only basic emotion containing a smile. The activation of AU12 (lip corners pulled up) resulting in a smile has been found to be very visually salient and the most important feature - maybe even the only necessity - for happiness recognition (Leppänen & Hietanen, 2007). Of the basic emotions, fear usually is the category hardest to recognise with the lowest recognition rates (e.g. Calvo & Lundqvist, 2008; Kirouac & Dore, 1985), which has also been found with the ADFES-BIV. The rank order of response latencies found with this study reflects the order of the accuracy rates and is in line with published research showing that negative emotions are more difficult to recognise than positive ones (Ekman & Friesen, 1976; McAndrew, 1986; Mufson & Nowicki, 1991; Ogawa & Suzuki, 1999; Rotter & Rotter, 1988), potentially due to overlapping AU activations between negative emotions. For example, an activation of AU4 (eyebrow lowered) resulting in a frown can be seen in an expression of anger as well as disgust making confusions between those emotions more likely. Logically, the more confusions occur between emotions, the lower the recognition rates for those emotions. The rank order of recognition for the emotion categories could hence be explained by the degree of featural overlap between the emotional facial expressions.

A further explanation for the particular rank order of recognition comes from research correlating the familiarity with particular emotions based on the frequency of encounters in daily life with recognition accuracy as well as response latencies and has found a strong positive relationship (Calvo et al., 2014). Precisely, Calvo et al. (2014) found happy faces to be experienced most often, followed by surprised, sad, angry, and disgusted faces, with fearful faces the least often experienced. This reflects the rank order of recognition within the current study and suggests that we are better at recognising the emotions that we encounter most frequently in our social interactions, simply because we are familiar with them. This familiarity would allow for a more accurate mental representation of features and configurations of the respective emotion facilitating its recognition (Calvo et al., 2014). Such a more accurate mental representation of an emotional facial expression should also lead to better recognition of some emotions at low expression intensity than others, which is indeed what has been found with the current study. For example, surprise was best recognised across intensity levels compared to the other emotion categories. That experience with an emotion influences recognition would suggest that individuals who encounter particular emotions more frequently than other individuals are also better at recognising the particular emotions than other individuals. It would be interesting to further examine individual differences in recognition of individual emotions from faces based on encountered frequency of the emotions. Of particular interest would be the role of intensity of expression within recognition. For example, if an emotion is better recognised by a person because it was encountered more often than by others, does the better recognition also apply to subtle displays of the respective emotion or only high intensity facial expressions?

In line with the results from the original ADFES (van der Schalk et al., 2011), the recognition rates of the complex emotions included in the current study led to lower recognition than the basic emotions, with contempt having been the hardest to recognise shown by the lowest recognition rates (see also Elfenbein & Ambady, 2002). It is possible that complex emotions are encountered less often than basic emotions and this diminished experience affects recognition. Future research should extend the study conducted by Calvo et al. (2014) to complex emotions to follow up this assumption. That emotions belonging to the category of basic emotions might be encountered more often and are generally better recognised than such considered complex emotions can be explained from an evolutionary perspective. It is claimed that we are innately set up to recognise the basic emotions, as this is functional for survival (Darwin, 1872; Ekman, 1992a). Complex emotions are not associated with securing survival and therefore maybe less often displayed and less well recognised.

Given that contempt has been suggested to also constitute a basic emotion (Ekman & Friesen, 1986), it should be more distinct and better recognised. The results of this study rather suggest that contempt is one of the complex emotions. This suggestion is furthered by the fact that the low contempt recognition rate in the current study is in line with the literature based on English-speaking subjects where contempt also was the least recognised of the emotions included in the research (e.g. Matsumoto & Ekman, 2004; Tracy & Robins, 2008; Wagner, 2000). The original ADFES was validated on a Dutch-speaking sample and could explain their better performance on contempt recognition (69%), because there is reason for the assumption that the concept 'contempt' entails slightly varying meaning between languages affecting recognition rates achieved by native English speakers (Matsumoto & Ekman, 2004). Matsumoto and Ekman (2004) showed experimentally that participants were able to match contempt situations, but failed when asked to freely label the emotion. They therefore suggested a probable cognitive conflict when searching for the label due to the linguistic closeness to the emotional state of content. Application of the ADFES-BIV in other languages than English with respective translations of the emotion terms would shed light on whether it is the category term 'contempt' that is problematic, or the videos are not properly representing contempt. It is possible that the prototypical display of contempt was not truly captured by the encoders in the videos or a different label (e.g. disapproval) would represent better the facial expression, which could be tested empirically within a freely labelling task.

Intensity level was shown to influence recognition rates and response latencies of the emotion categories differently. However, the general pattern was that the emotional expressions were easier to recognise when the intensity level was higher (except for sadness where the intermediate and high intensity expressions were equally recognised). The difference between the emotion categories was that the increase of accuracy of response from one intensity level to the next varied between emotions. For example, the increase in accuracy from happiness expressed at low intensity compared to intermediate intensity was 23%, but the increase for surprise was only 2% from low to intermediate intensity. It seems that happiness is more ambiguous at low expression intensity. The importance of a salient smile in happiness for its recognition (Calvo & Nummenmaa, 2011) might explain the lower recognition at low expression intensity where only a little smile is visible in the stimuli. The very good recognition of surprise even at low intensity might be rooted in the importance of surprise recognition. Surprise is just the initial reaction to an unexpected stimulus and then turns into one of the other basic emotions (Ekman & Friesen, 1975), for example happiness or fear. This is also the reason why it has been equated with the startle response by some (Stein, Hernandez, & Trabasso, 2008), giving surprise substantial signalling/alerting character and making it necessary to be recognised easily to prepare for action. Alternatively, surprise could therefore be deemed the most basic of the basic emotions, making the recognition less dependent on expression intensity.

Sex differences.

As hypothesised, sex differences in accuracy of response and response latencies were found with the current study, with females generally outperforming males at facial emotion recognition. This result is in line with previous studies and meta-analyses of sex differences in facial emotion recognition (e.g. JA Hall, 1978, 1990; JA Hall & Matsumoto, 2004; JK Hall et al., 2010; Kirouac & Dore, 1985; Kret & De Gelder, 2012; Lee et al., 2013), but not with other reports (e.g. Grimshaw et al., 2004; Rahman et al., 2004). The reported effect sizes for sex differences in facial emotion recognition are rather small, which makes a larger sample size necessary as well as a sensitive task to reveal sex differences and could explain the non-significant findings for sex differences in facial emotion recognition in some studies. The results from the current study support the notion that females outperform males in emotion processing, meaning females achieve higher accuracy of response and are also quicker at recognising emotions from faces (e.g. Collignon et al., 2010; JK Hall et al., 2010; Rahman et al., 2004). It can be concluded that females indeed recognise emotions from faces faster and more correct than males.

It was hypothesised that females would outperform males at facial emotion recognition particularly for emotions that are more difficult to recognise. However, the results from the current study did not show that males and females differed in which emotions showed sex effects based on accuracy of response rates or response latencies. A general female advantage over males in response latencies was found with the current study, not influenced by emotion, intensity, or an interaction of both. That emotion did not influence the female advantage is in line with JK Hall et al. (2010), Rahman et al. (2004), and Lee et al. (2013), whereas Hampson et al. (2006) found the female advantage to be most prominent for negative emotions (although an interaction was not reported in the paper). That is, females recognise emotions faster than males from faces. Previous reports have been mixed in that the female advantage based on accuracy of response was (e.g. Kirouac & Dore, 1985) and was not (e.g. JK Hall et al., 2010) found to be affected by emotion. Sample size might be a determining factor in regards to a significant interaction effect of group and emotion. The studies that did find such an effect included large samples, e.g. 300 participants in the study by Kirouac and Dore (1985), as opposed to 50 participants in the study by JK Hall et al. (2010) and 92 in the current study. If a very large sample size is needed to obtain significant results, this indicates that the underlying effect is negligible. The advantage of the current study is that it included not only the six basic emotions but also three complex emotions. Since a wider range of emotion categories was included in the task, the results suggest that females are consistently better at recognising emotions from faces than males across emotion categories.

Taking into account intensity of expression shows more fine-graded differences between the sexes in facial emotion recognition. Based on the reports by Hoffmann et al. (2010), it was hypothesised that females would outperform males only at recognition of subtle emotional expressions, but not high intensity expressions. Hoffmann et al. (2010) found that increasing the intensity of expression increased recognition in both sexes, although there was a steeper decline in accuracy of response for males with decreasing expression intensity than for females, which contributed to the significant differences at low expression intensity in their study. In contrast, in the current study, the female facial emotion recognition advantage was given across intensity levels and the advantage was greatest at intermediate intensity of expression. It is possible that the task difficulty was generally higher with the ADFES-BIV than within the study by Hoffmann et al. (2010). The results can be interpreted the way that when the task is very easy or very difficult, then no group differences emerge. For the ADFES-BIV that means, at low expression intensity the task was rather hard to accomplish (but not too hard), and at high expression intensity rather easy (but not too easy). Therefore, sex differences in facial emotion recognition appeared throughout the intensity levels, yet were greatest at intermediate expression intensity.

The influence of difficulty on results also applies to sex differences in the recognition of individual emotions at the varying intensity levels of expression. The current study found the female advantage over males in facial emotion recognition to be strongest at intermediate expression intensity, as for most emotion categories significant differences were found at intermediate expression intensity (not for surprise and pride). This is because subtle emotional expressions of all emotion categories are more difficult to recognise and high intensity emotional expressions easier to recognise. As a result, expressions at intermediate expression intensity have the most discriminative ability. As Hoffmann et al. (2010) pointed out, easy to recognise stimuli have less discriminative ability, which is why sex differences are more likely to show based on stimuli of intermediate difficulty. Emotions that are easily recognised (e.g. surprise) or difficult to recognise (e.g. pride) across expression intensity levels do not have much discriminative power. This is because in the former no high recognition ability is required, so both sexes manage well, and in the latter case high recognition ability is required and both groups struggle a little. In both cases it is less likely to find significant group differences in facial emotion recognition. In line with this explanation, the female advantage in the current study was already apparent in the low intensity expressions for emotions that were easier to recognise (embarrassment and sadness). Males needed higher expression intensity to achieve accuracy of response comparable to females, explaining why significant differences were found at low but not high expression intensity. Females outperformed males on all intensity levels of anger and disgust, emotions with intermediate recognition difficulty. Differences showed in the high intensity for emotions that were harder to recognise, like contempt, as both sexes had difficulty with those subtle expressions, but females performed well at high intensity. A pattern of group differences can be expected in order from no differences to differences at low intensity to differences at intermediate intensity to differences at high intensity to no differences as difficulty to recognise the individual emotion increases. A visualisation of this pattern can be found in Appendix E. It can be concluded that females are generally better at recognising facial expressions; they need less intensity of emotional cues in the expression than males and difficulty to recognise an emotion modulates where the female advantage becomes most prominent (explanations follow below).

While difficulty determines when and where the female advantage in facial emotion recognition shows, it does not explain why females outperform males. It is possible that females process emotional faces in a different manner than males. Hampson et al. (2006) tested the possibility that females are simply faster in processing faces (perceptual speed) and found this not to be the case and the female advantage to be specific to emotion recognition (no female advantage was found for face identity recognition). A female advantage has also been reported by Hoffmann et al. (2010) who used short exposure times to emotional content (300ms) and also in the current study exposure time to emotional context was shorter than 1s (varied between intensities). It has to be noted that males and females did not differ in processing of neutral faces within the current study. Those results lead to suggest that whenever quick judgements about facial emotion are necessary, females outperform males. This assumption also fits in well with the literature reporting fast and automatic processing of facial emotion in females (e.g. Hall & Matsumoto, 2004), and fast and automatic processing specific to facial emotion (e.g. Donges et al., 2012). Fast and automatic processing of facial emotion might be innate as proposed by the primary caretaker hypothesis (Babchuk et al., 1985). In line with the reports by Hampson et al. (2006), the current study found support for the attachment promotion hypothesis and the fitness threat hypothesis, since females recognised positive (happiness) and negative emotions (anger, disgust, fear, sadness, contempt) better than males. However, the female recognition advantage might also have been acquired as part of females' socialisation. As mentioned above, research by Calvo et al. (2014) showed that recognition and familiarity with facial emotional expressions are linked. It is therefore possible, that females have developed their greater emotion processing abilities due to more exposure to emotional displays.

Research from neuroscience has shown that females differ from males in their pattern of brain activity during facial emotion recognition. The right inferior frontal cortex was found to be differentially activated in females but not in males during facial emotion recognition (Schulte-Rüther, Markowitsch, Shah, Fink, & Piefke, 2008). This region is thought to host mirror neurons, the activation of which might facilitate facial emotion recognition based on simulation of the observed facial emotional expression (see Chapter 2).

Interestingly, in males the left temporoparietal junction showed differential activation and has been related to the distinction between self and others. Schulte-Rüther et al. (2008) hence suggested that males and females differ in the strategies applied when evaluating emotions. It is possible that females recognise emotions from faces based on a simulation approach like through shared affect as suggested by the reverse simulation model (Goldman & Sripada, 2005) and males rely more on information-based strategies fitting in with theory-theory (Morton, 1980). Indeed, females have been found to pay more attention to the eyes during facial emotion recognition (JK Hall et al., 2010) and eye contact has been theorised to be central for simulation processes (Niedenthal et al., 2010). An interesting postulation has been made by Rahman et al. (2004) saying that males and females might have differing objective impressions as to when a happy or sad expression is present or might differ in the perception of how an emotion is expressed, rather than that males and females differ in the capability to discriminate emotions. This postulation fits with reports by Lee et al. (2002) that males and females differ in the areas of brain activation during emotion recognition of happy and sad facial expressions and should gain attention in future research.

Limitations.

The varying durations of emotional content visible between the intensity levels of the ADFES-BIV could be seen as limitation in terms of standardisation. However, they could also be interpreted as a further indicator of ecological validity, as the point of onset always precedes the reaching of maximum expression intensity. It is virtually impossible to create stimuli of facial expression at varying intensities and have the emotional content visible for the exact same duration between the intensities without manipulating the expression dynamics. Given the anatomical underpinnings of muscular movements, expressing low intensity facial emotional expressions should take less time than high intensity ones, since a more intense muscle contraction should take more time. Except if it was the case that we speed up our facial expression when expressing full intensity and slow down for subtle expressions which seems unlikely, but research on that has yet to be undertaken. Further support comes from the DaFEx video database (Battocchi et al., 2005) where professional actors were specifically instructed to express low, intermediate, and high intensity emotional facial expressions. The resulting videos were shorter in duration in the low expression intensity condition than the intermediate intensity ones and the intermediate intensity videos were shorter than the high intensity videos, just as in the current study. It can therefore be concluded that even though the current videos were edited, they still resemble a natural occurrence for posed expressions – the intended purpose of the stimuli.

A limitation of the stimuli is that the videos were edited instead of instructing encoders to express low, intermediate, and high emotional facial expressions. However, this does not propose a strong limitation, since muscles can only do specific movements, which are catalogued in the FACS (Ekman & Friesen, 1978), it is only the intensity of muscle contraction that changes. For example, when making a sad facial expression as intense as possible, it always starts out as a subtle expression increasing to full intensity. Hence, it is legitimate to extract sequences and still being able to claim to have varying intensity levels. Yet, future research could instruct encoders to pose varying intensities of facial expression and FACS code them to verify the present results. Future research should also aim to produce videos of varying intensity portraying individuals expressing emotions facially that have been elicited in those individuals, i.e. truly felt emotions, to further increase ecological validity. Subjective ratings alongside FACS coding would be necessary to assure the elicited emotions equal the target emotion.

As part of the study presented here, the three levels of expression intensity that have been created were validated and differences in accuracy of response and response times between the expression intensity levels were found, with more correct and faster responses to high intensity expressions than low intensity expressions. This finding was explained by the intensity of expression itself. However, there were differences between the intensity levels in display time of the emotional expressions seen by participants. In the low intensity videos of the ADFES-BIV the emotional expression was visible for less time than in the intermediate and high intensity videos, and the intermediate intensity videos had the expression displayed for less time than the high intensity videos. The resulting differences in processing time could be underlying the results, rather than the intensity of the facial expressions. Since the results from the current study cannot shed light on this, a further study needed to be conducted investigating the potential influence of exposure time on accuracy of response and response latencies within facial emotion recognition. With the aim to demonstrate that the results from the current study were not affected by exposure time, a further study was conducted controlling for exposure time. This study is presented in the following chapter.

CHAPTER 7—Validation of the ADFES-BIV intensity levels

Introduction

The study presented in Chapter 6 showed validation of the ADFES-BIV levels of expression intensity based on accuracy rates and response latencies. One potential explanation for the found differences in accuracy of response and response latencies between the intensity levels is that the differences in accuracy of response and response times are based on exposure time to emotional content in the videos, rather than intensity of expression, as discussed in Chapter 6. It was therefore necessary to confirm the results presented in Chapter 6 with stimuli of constant exposure time to the faces themselves as well as to the emotional content. If the amount of time the facial emotional expression was displayed across intensity levels was causing differences in accuracy rates and response latencies, these differences across intensity levels should be lost with this variation of the videos as the display times were equated. Instead, if it is the degree of intensity that is important rather than the amount of time the expression is displayed, then differences in accuracy rates and response latencies should be evident across the different intensity levels similar to the findings from the study presented in Chapter 6.

Aims and hypotheses

Aim of the current study was to validate the intensity levels of the ADFES-BIV under controlled timings by testing the relationship of exposure time to emotional content in faces (based on recognition rates and response latencies).

It was hypothesised that the same pattern of accuracy of response and response times would be found as in Chapter 6. That is, highest accuracy rates and fastest responses to the high intensity videos, lowest accuracy rates and slowest responses to the low intensity videos, and accuracy rates and response times to the intermediate intensity videos in between.

Method

Participants.

Thirty individuals (15 females, 15 males) were recruited from the University of Bath community. Participants were PhD students (*n* = 16), Undergraduate students (*n* = 10), Masters' students (*n* = 1), and staff (*n* = 3) from diverse departments including Psychology (*n* = 15), Health (*n* = 3), Mechanical Engineering (*n* = 9), Computer Sciences (*n* = 2), and Physics (*n* = 1). All participants had normal or corrected-to-normal vision to assure the ability to perceive the presented emotion cues. Participants were compensated with £5 or gained course credit (Psychology students) for participation. The age ranged from 18 to 60 years (*M* = 27.7, *SD* = 9.9). One person indicated having a diagnosis of an Anxiety Disorder. Since the data for this participant was comparable to the other participants in terms of their accuracy rates and response latencies, their data remained included in the analysis. Due to a technical fault response time data were only available for 29 participants (-1 male). Ethical approval for this study was given by the University of Bath Psychology Ethics Committee (Ref 13-137).

Stimuli.

To keep the exposure time constant, a first-last frame approach was chosen for developing the stimuli of the ADFES-BIV with controlled timings. That is, only the first and last frame of each of the ADFES-BIV videos are shown. As a consequence of this approach, the progression of the facial expression is not visible and the temporal characteristics of the individual emotional expressions are lost. However, some extent of motion remains due to the change from neutral to an emotional facial expression. Since temporal characteristics are argued to be part of our emotion representations (Sato & Yoshikawa, 2004) and therefore aid recognition (Bould et al., 2008; Bould & Morris, 2008; Kamachi et al., 2013), the first-last frame approach generally leads to lower accuracy rates than complete dynamic sequences (see Bould et al., 2008, Bould & Morris, 2008).

From the 360 videos of the ADFES-BIV the first and last frame were extracted for each video. Video sequences were created with the first frame (neutral expression) being presented 13 times and then the last frame (emotional expression) for the same amount of repetitions with a speed of 25 frames/sec. It was chosen to apply the exact same characteristics to the first-last frame stimuli as to the ADFES-BIV videos, i.e. equal amount of frames (26) and frame rate (25 frames/sec) as well as starting with a neutral image and ending with an emotional expression. The duration of the first-last frame videos was 1040ms per video equivalent to the ADFES-BIV videos. Thereby, highly standardised videos in regards to timings across the levels of expression intensity were created.

Procedure and task.

The same task and procedure were applied as for the studies in Chapters 4 and 6.

Data preparation and analysis.

Accuracy rates were calculated for each level of *intensity*. Inspection of the Shapiro-Wilk statistics indicated that the accuracy rates for the intensity levels followed a normal distribution (low: D = .98, df = 30, p = .698; intermediate: D = .94, df = 30, p = .115; high: D = .95, df = 30, p = .201). No extreme values were identified inspecting boxplots. A repeated measures ANOVA with *intensity* as within-subject factor was hence conducted to test if the intensity levels differed significantly in accuracy of response.

Response times were calculated for the *intensity* levels based on correct trials only. Inspection of the Shapiro-Wilk statistics revealed the response time data for the intensity levels were significantly different from a normal distribution (low: D = .93, df = 29, p= .048; intermediate: D = .90, df = 29, p = .009; high: D = .93, df = 29, p = .062). The data were therefore normalised using log transformation (low: D = .99, df = 29, p = .994; intermediate: D = .97, df = 29, p = .656; high: D = .98, df = 29, p = .764). No extreme values were identified inspecting boxplots. A repeated measures ANOVA with *intensity* as within-subject factor was conducted to test if the intensity levels (correct trials only) differed significantly in response time.

Results

Accuracy of response.

The repeated measures ANOVA revealed a significant main effect of *intensity* ($F(2,58) = 252.70$, $p < .001$, partial $\eta^2 = .897$, power = 1.00). Taken together, a mean accuracy of response of .51 ($SD = .09$) was achieved for the low intensity videos. The mean accuracy of response for the intermediate intensity was .66 ($SD = .12$) and .74 for the high intensity level ($SD = .10$). Pairwise comparisons showed the three intensity levels were all significantly different from each other (p's < .001); see Figure 18.

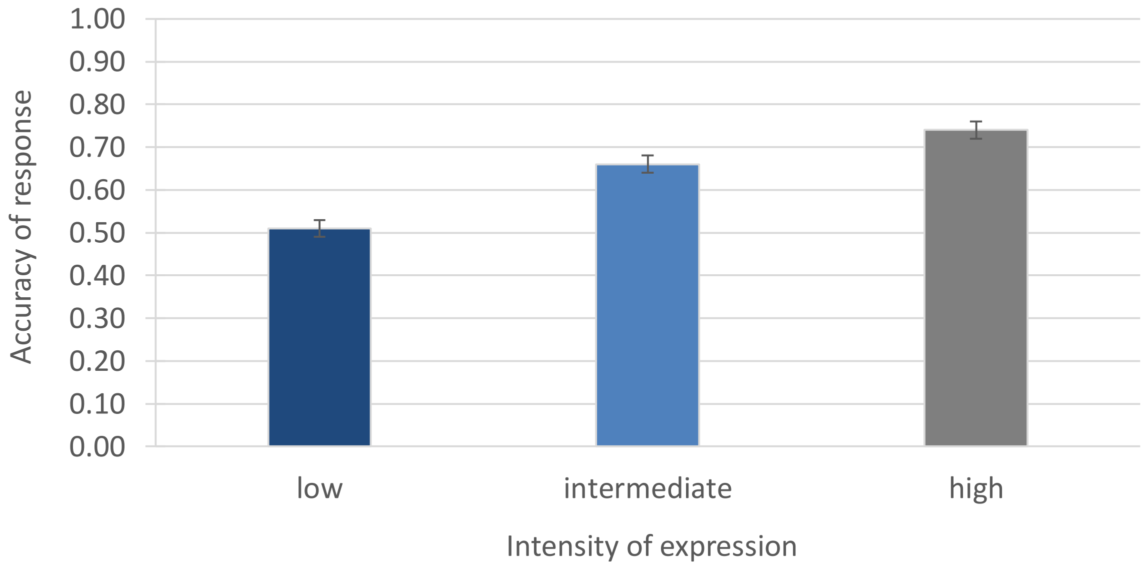


Figure 18. Accuracy of response for the three intensity levels of the ADFES-BIV first-last frame stimuli. 1 = 100% accuracy of response. Error bars represent standard errors of the means.

Response times.

The repeated measures ANOVA with *intensity* as within-subject factor revealed a significant main effect of *intensity* ($F(2, 56) = 4.28$, $p = .019$, partial $\eta^2 = .133$, power = .724). Pairwise comparisons showed that the videos from the high intensity level ($M = 1054$ms, $SD = 399$) were recognised significantly faster than the videos from the intermediate intensity level ($M = 1137$ms, $SD = 491$, $p = .014$). Responses did not occur significantly faster for the intermediate intensity level than to the low intensity level ($M = 1171$ms, $SD = 452$, $p = .480$).

Responses to the high intensity level were significantly faster than to the low intensity level (p = .011) (see Figure 19).

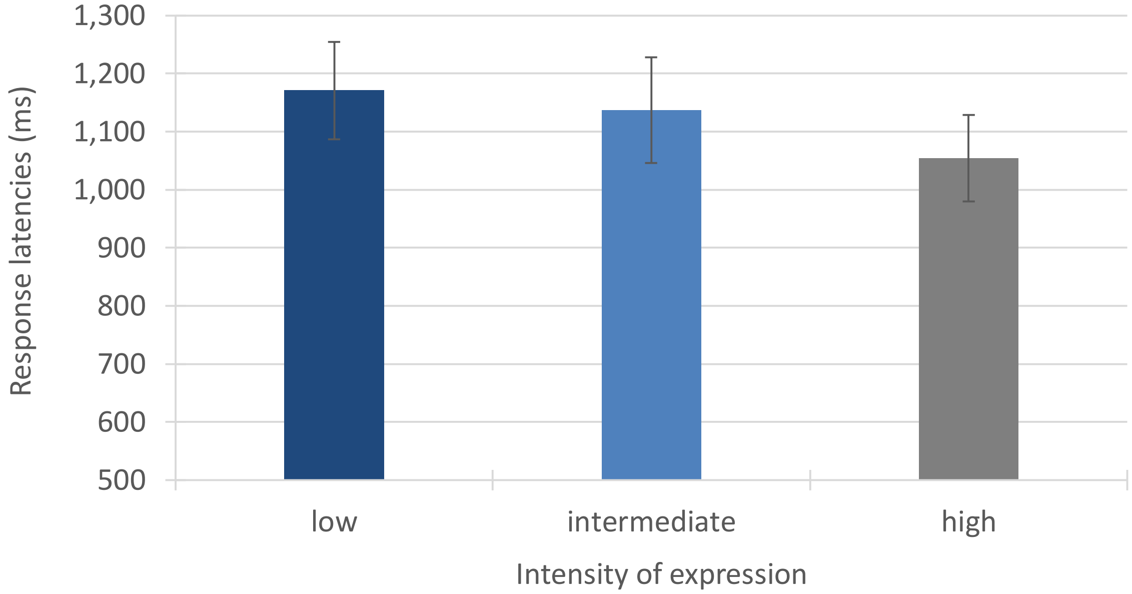


Figure 19. Response times (ms) for the intensity levels of the ADFES-BIV first-last frame stimuli. Error bars represent standard errors of the means.

Discussion

The main aim of the current study was to demonstrate that the retrieved differences in accuracy rates and response latencies between the intensity levels of the ADFES-BIV that are presented in Chapter 6 were the result of the intensity variations and not due to the varying exposure time to emotional content between the intensity levels. This demonstration was intended to further validate the ADFES-BIV intensity levels. The expected pattern of accuracy of response was found for the three intensity levels of expression, with lower accuracy of response for the low intensity level compared to the intermediate intensity level, which in turn had lower accuracy of response than the high intensity level. The pattern of results was not as clear for the response latencies. That is, responding occurred not significantly faster to intermediate than low intensity expressions, although responding occurred significantly faster to high than intermediate intensity expressions. Results show that exposure time is less relevant for facial emotion recognition than the intensity of facial expressions. Discussion of these findings follows below.

The display time of emotional content was kept constant in the first-last frame videos to investigate the influence of intensity of expression on facial emotion recognition. If the display time were the factor to modulate the accuracy rates for these videos, the intensity levels would not have been significantly different from each other, as the display time of emotional (and neutral) content was exactly the same for the different intensity level. The accuracy rates from the high intensity facial expressions were significantly higher than from the intermediate intensity expressions. The accuracy rates from the intermediate intensity expressions were significantly higher than from the low intensity facial expressions. Therefore, it can be concluded that even though the exposure time was kept constant, the low intensity category was still more difficult to correctly recognise than the intermediate one and the high intensity expressions were easiest to recognise. This demonstrates that it is the intensity of expression, and not display and processing time, which influences recognition accuracy from the videos.

A similar pattern of results to the accuracy rates was found for the response times. The mean response time for the low intensity stimuli was 117ms longer than for the high intensity stimuli, which is in line with the prediction. This means that even when the same processing time is allowed for the different intensities, responses occur slower for low intensity expressions than for high intensity expressions. Just as for accuracy of response, the expression intensity is of higher relevance for responding times than the display time. However, the intermediate intensity level was not different to the low intensity level, although it was expected that responding would occur faster to intermediate than low intensity expressions. A potential explanation is, that reducing the number of frames to one frame showing a neutral face and one frame containing emotional information, simply did not offer enough information to facilitate responding as opposed to when the whole progression of the expression is shown as in the videos of the ADFES-BIV with varying exposure time. It can be concluded that the varying exposure times to emotional content between the intensity levels have little influence on response times. The results support the distinction of low, intermediate, and high intensity, since the categories were found to differ from each other in accuracy of response and mostly in response latencies and validates the intensity levels of the ADFES-BIV as distinct categories.

The studies presented in this dissertation so far were concerned with developing and validating a video-based stimulus set of facial emotional expressions with different intensities of expression, the ADFES-BIV. By demonstrating with the study from the current chapter that the varying levels of expression intensity were not affected by the differences in exposure time to emotional content and that it is the intensity of expression shaping the difficulty of recognition, the stimulus set was effectively validated. Since the ADFES-BIV proved suitable for assessment of group differences in facial emotion recognition ability in the general population (Chapter 6) and the facial emotion recognition task seemed manageable by individuals with high-functioning ASD (Study 2 in Chapter 4), the ADFES-BIV was next applied in a study assessing facial emotion recognition in high-functioning autism compared to controls. The study is presented in the following chapter.

CHAPTER 8—Facial emotion recognition in high-functioning autism

Introduction

One of the main characteristics of ASD is difficulties in communication skills and social functioning which includes non-verbal communication (APA, 2013). Since expressing emotions facially is one form of non-verbal communication, it was derived that individuals with ASD might have difficulties inferring emotional states from faces. This assumption has led to a plethora of research investigating the impairments surrounding non-verbal communication in ASD with facial emotion recognition of particular interest. In Chapter 1, it was discussed that individuals with ASD often show deficits in facial emotion recognition when compared to typical controls, evidenced in literature reviews (Harms et al., 2010; Nuske et al., 2013) and meta-analyses (Lozier, Vanmeter, & Marsh, 2014; Uljarevic & Hamilton, 2013). Nonetheless, these literature reviews and meta-analyses also show that not all behavioural studies on facial emotion recognition in ASD show impairments when compared to typical individuals. According to Harms et al. (2010) almost half of the behavioural studies on facial emotion recognition in ASD do not report a recognition deficit in ASD. Variations in applied methodology account for these inconsistent findings to some extent. Since research mainly focuses on high intensity facial emotional expressions, but subtler displays of facial expression are part of social interactions, research is needed including subtler displays of facial emotional expression to allow for a more reliable assessment of facial emotion recognition.

Using stimuli with low as well as high intensities of emotional expression might produce more variability in responses and better allow for group comparisons. There have only been a few studies published that applied varying levels of expression intensity. What these studies have in common is that they all report a facial emotion recognition deficit in ASD. For example, Mazefsky and Oswald (2007) investigated recognition of low and high expression intensity of anger, happiness, fear, and sadness in children with high-functioning ASD based on static images. No typical control group was included in the study, instead the standardisation sample from the applied stimuli's manual was used for comparison. A significant difference was found for the overall accuracy of response between the high-functioning ASD group and the standardisation sample. Since the manual underlying the applied stimuli did not report standardised data by intensity, an intensity-specific comparison could not be undertaken by Mazefsky and Oswald (2007), neither did the authors test for group differences on the level of emotion categories. Nonetheless, this study shows that when including low intensity expressions next to high intensity facial expressions, significant group differences emerge.

However, the facial emotion recognition deficit in ASD seems to be more likely to appear in adulthood. Rump, Giovannelli, Minshew, and Strauss (2009) compared facial emotion recognition ability of three age groups of individuals with an ASD to matched controls; children, adolescents, and adults. Since facial expressions are dynamic and therefore to increase ecological validity further, Rump et al. (2009) used video recordings of the six basic emotions and investigated facial emotion recognition with varying intensity of expression. The videos were manipulated to an exposure time of 500ms for each video, across all intensities (- the authors did not state whether this was achieved by elimination of frames or alterations of the display frame rate). Prominent differences to controls were found only in adults and for the emotions anger, disgust, fear, and surprise; a ceiling effect was present for happiness and floor for sadness, which is why group comparisons could not be conducted for happiness and sadness. The found recognition deficit only in adults with ASD suggests that the development of facial emotion recognition continues into adulthood to a lesser extent in individuals with ASD than in typical individuals. The study by Rump et al. (2009) further suggests that the facial emotion recognition deficit in ASD is not emotion-specific.

Even though the presented studies show that it is useful to include varying intensities of facial emotional expression to reveal an overall facial emotion recognition deficit in ASD compared to controls, investigating facial emotion recognition at the level of emotion categories and intensity level per emotion can provide further interesting information as to where recognition deficits are located. Law Smith, Montagne, Perrett, Gill, and Gallagher (2010) conducted a study applying morphed dynamic stimuli of low, medium, and high expression intensity of the six basic emotions and compared male adolescents with ASD to age, sex, and IQ matched controls regarding the recognition accuracy. The authors found the ASD group to be impaired at recognising the emotions disgust, anger, and surprise. The ASD group was outperformed by the controls at each of the expression intensity levels, but the group difference in accuracy of response was greatest at the medium intensity level. This finding is interesting, as one could have assumed that individuals with ASD have the most trouble with recognising subtle facial emotional expressions based on their difficulties with interpreting facial cues leading to greatest group differences at the low intensity level. This finding is further interesting, as it follows the same pattern to the results from the comparison of typical males to typical females (study presented in Chapter 6) and might suggest that facial emotion recognition impairments show the same pattern across populations in regards to intensity of expression, only differing in the magnitude of differences. Law Smith et al. (2010) also found the level of expression intensity at which impairments were identified in the ASD group to vary between emotion categories. The recognition of disgust was impaired in ASD at all intensity levels and recognition of surprise and anger was comparable to controls only at the high expression intensity level. Inclusion of only high intensity facial expressions would have led to different findings and the false assumption that only disgust recognition is impaired in ASD. This study demonstrates the importance to include varying expression intensities and to include all levels of categories in the analysis to learn more about the specifics of the facial emotion recognition deficit in ASD.

Although the general consensus is that individuals with ASD have deficits in facial emotion recognition, much less is known about the underlying mechanisms bringing about the deficits. Since confusions (i.e. false responses or errors) within a facial emotion recognition task are essentially what causes lowered accuracy rates, the investigation of confusions can provide substantial information about the processes manifesting in deficits in facial emotion recognition. Confusions of an emotion with 'neutral' give information about the sensitivity to detect emotional content in the face. Confusions of one emotion category with another emotion give information about the ability to differentiate between facial emotional expressions (specificity). The latter confusion is naturally more likely to occur for emotions that overlap featurally, such as anger and disgust (frown) or fear and surprise (wide open eyes). Focussing on one single feature makes confusions more likely for featurally similar emotional facial expressions. Whereas there is a wealth of literature on facial emotion recognition in ASD reporting accuracy rates, confusions are rarely reported despite their informative nature.

However, Humphreys, Minshew, Leonard, and Behrmann (2007) did report confusions in a confusion matrix and showed that individuals with ASD tended to confuse facial expressions of disgust with anger and fear with surprise – emotional expressions that are featurally similar. These findings fit in with the weak central coherence theory (Frith, 1989/2003) according to which individuals with ASD focus more on single facial features during decoding of facial expressions than typical individuals. As said above, the focus on single features makes confusions more likely. Nevertheless, the particular confusions reported by Humphreys et al. (2007) are 'normal', as they are often seen in typical individuals (e.g. Recio et al., 2013). This is backed up by Jones et al. (2011) who also presented confusion matrices and stated that the ASD group showed the same pattern of confusions as the controls. When investigating confusions, additional statistical comparisons of the confusions made by individuals with ASD and typical individuals would provide useful information. The arising and important question is whether individuals with ASD make significantly more such 'normal' confusions than typical individuals.

To my awareness, there is only one study that carried out statistical comparisons for the confusions made by the ASD group and controls. Eack, Mazefsky, and Minshew (2015) used static images of actors portraying happiness, fear, sadness, anger, and neutral to investigate facial emotion recognition. They found overall lower accuracy rates for the ASD group compared to controls and the impairment to be greatest at recognition of neutral and happiness. The finding of the greatest deficit for recognition of happiness is rather unusual, as the majority of the literature reports no impairment for happiness (literature review by Harms et al., 2010). When investigating the confusions, Eack et al. (2015) found the individuals with ASD to misinterpret neutral faces as displaying an emotion (mostly sadness and anger) significantly more often than controls, indicating a negativity bias. They also found individuals with ASD to perceive emotional faces (happiness and sadness) as neutral, indicating a lowered sensitivity to detect emotional content from faces. The found misattribution of sadness as fear that occurred significantly more often in ASD than controls indicates an issue of specificity based on a lack of the ability to discriminate between emotional expressions. Eack et al. (2015) demonstrate with their study that it is imperative to investigate confusions to learn more about the specifics of the deficits of individuals with ASD in facial emotion recognition.

Even though assessment of facial emotion recognition at varying levels of intensity of expression allows for a more reliable and fine-graded investigation of the ability to recognise emotions from faces and increases ecological validity, only few studies are published on facial emotion recognition including varying intensities of expression. Of those studies, not all investigated all six basic emotions and used either static or manipulated dynamic stimuli. The need for facial emotion recognition investigations in ASD based on video recordings of varying intensities of facial emotion including a wider range of emotions is thus highlighted. With only a limited amount of studies presenting confusions and only one published study that statistically compared confusions within facial emotion recognition in ASD to controls, the need for more such investigations in emphasised with the purpose to advance our understanding of facial emotion recognition deficits in ASD. To address these needs, the current study applied the ADFES-BIV comparing adolescents/young adults with a diagnosis of high-functioning autism to matched controls. The wider range of emotions and inclusion of varying intensities of expression allowed for emotion-specific and intensity-specific investigation of group differences and confusions.

Aims and hypotheses

The specific aims of the study were:

(1) to compare adolescents/young adults with ASD to controls on facial emotion recognition ability based on accuracy of response,
(2) to compare the confusions made by the ASD group statistically to those made by typical individuals, and
(3) to investigate the confusions made by the ASD group in regards to sensitivity and specificity in facial emotion recognition.

Based on the outlined literature, it was expected that

(1) individuals with ASD would show an overall deficit in facial emotion recognition compared to controls,
(2) the deficit in facial emotion recognition in ASD would be specific to certain emotions (negative emotions) and not others (positive emotions),

(3) the deficit in facial emotion recognition in ASD would show across intensity levels,
(4) the pattern reported by Law Smith et al. (2010) would be replicated when applying video stimuli, i.e. the recognition of the emotions would be influenced differently by the level of intensity in the two groups with deficits in recognition of surprise and anger only at the lower intensities but impairment in recognition of disgust across intensity levels, and
(5) the ASD group would show significantly greater impairments in specificity based on the weak central coherence theory (Frith, 1989/2003), i.e. be more likely than typical individuals to confuse emotions that overlap featurally in their facial expression (e.g. anger and disgust, fear and surprise), and that individuals with ASD would show greater impairments in sensitivity, i.e. would have problems with detecting emotions from faces evidenced by significant confusions with neutral.

Method

Participants.

The sample consisted of 12 individuals aged 16-19 with high-functioning autism (9 male, 2 female, 1 transgender with female sex and male gender) (M = 17.3, SD = .75) and 12 controls aged 16-17 matched on sex (9 male, 3 female) and age (M = 16.9, SD = .29). The mean age difference between the two groups was not significant as indicated by an independent-samples t-test ($t(14.16) = -1.43$, $p = .174$). All participants were British and had normal or corrected-to-normal vision to avoid any group differences to be the result of visual impairments. All participants were of a similar intelligence level, since they were all A-level students in the process of entering or applying to university. The ASD sample was recruited and tested at an Autism Summer School run at the University of Bath for individuals with ASD close to entering university. ASD diagnoses were given by clinical professionals at some stage in the participants' life and the ASD participants brought a copy of diagnosis to the Autism Summer School. Controls were also students in the process of applying to university and were recruited tested at a University of Bath Open Day. One ASD participant reported to have received training in facial emotion recognition before and one participant in each group reported to have participated in a facial emotion recognition study before. Those two participants scored in the middle range of accuracy of response and therefore training effects are not expected to have affected the results presented here. Ethical approval for this study was given by the Department of Psychology Ethics Committee (REF: 13-137).

Questionnaires.

The AQ (Baron-Cohen et al., 2011) was used to screen for an ASD in the ASD group and the absence of an ASD in the control group. The AQ is widely used self-report tool for assessing autism traits (Ruzich et al., 2015) and validated (Baron-Cohen et al., 2001). The AQ comprises of 50 items each rated on a 4-point Likert scale. 4-point Likert scale, with answer categories “1 = definitely agree”; “1 = slightly agree”; “0 = slightly disagree” and “0 = definitely disagree” resulting in a minimum of 0 and a maximum overall score of 50. The scoring is reversed for the 24 items where ‘disagree’ is indicative of autistic traits. Sample items are: “I find myself drawn more strongly to people than to things” and “I find it easy to work out what someone is thinking or feeling just by looking at their face”. The AQ has satisfactory internal consistency with Cronbach's α varying from 0.63 to 0.78 (Baron-Cohen et al., 2001; Kurita, Koyama, & Osada, 2005). In a systematic review, Ruzich et al. (2015) found the AQ to be a sensitive measure for the assessment of autism traits in the non-clinical population and extending into clinical ASD. A cut-off of 26 in the AQ sum score for a potential ASD was applied (Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005). Next to the self-report, parents-report via the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) was used to assess autistic behaviour. The SCQ is a 40-item questionnaire of yes-or-no questions (scored 1 and 0 respectively) and the sum score of the SCQ ranges from 0 to 39. The recommended cut-off score for a potential ASD is 15 with a sensitivity of .85 and specificity of .75. The SCQ has excellent internal consistency (.90) making the SCQ an effective screening measure for ASD. All participants in the ASD group in the current study were diagnosed at some stage in their lives by a clinical professional. For attendance at the Autism Summer School, participants had to bring a copy of their diagnosis. The ASD participants within this study were hence treated as having ASD.

Facial emotion recognition task.

The ADFES-BIV videos as used in the study presented in Chapter 5 were used. The task was identical to what is described in Chapter 4 and used in Chapter 6. The experiment was set up to record the responses (emotion label) given by participants and the correctness of the response per trial (1 = correct, 0 = incorrect).

Procedure.

The procedure was similar to the description in Chapter 5, except that more participants were tested at once based on the time constraints. Testing sessions were conducted with one to seven participants at a time. At the end of the testing session participants were fully debriefed and the control group participants were given £5 for participation.

Data preparation and data analysis.

As discussed in Chapter 5, GLMM is an appropriate statistical approach for accuracy of response data and was conducted again. Since the only assumption for GLMM is that there are no extreme values, the data was inspected with boxplots and no extreme values were identified. The analysis was therefore conducted with the full, unaltered data set.

The data structure for the GLMM was defined with ‘subject’ and ‘group’ as subject specifications, ‘emotions’ and ‘intensity’ as repeated statements with diagonal covariance structure. ‘Subject’ was included as random factor including the intercept. The fixed factors were ‘intensity’, ‘emotions’, ‘emotions*intensity’, ‘group’, ‘group*emotions’, ‘group*intensities’, and ‘group*emotions*intensities’. A binomial distribution with logit link function were specified, as appropriate for proportion data. Degrees of freedom were estimated by the residual method. Robust estimation of fixed effects and coefficients was applied to account for potential model violations. Simple contrasts were requested to compare the groups and pairwise comparisons were requested to compare the emotions and intensities among each other. Sequential Bonferroni corrections were applied to correct the *p*-value for multiple comparisons within the contrasts. Neutral was not included in the main analysis, since there are no intensities of expression for this category. However the ASD group was compared to controls on the recognition of neutral faces based on a Mann-

Whitney *U*-test, since the variable neutral was non-normal in the ASD group ($W(12) = .84$, $p = .027$), with the purpose to test for perceptual differences of faces.

Confusion matrices were created in E-Prime for both groups separately. The matrices include the responses given to the emotions displayed expressed in percentages for each emotion separate for the low and high intensity expressions. In SPSS, variables were created for all the confusions that occurred (e.g. anger at low intensity as disgust). The confusions that occurred more often than by the 10% chance level of responding based on visual inspection of the confusion matrices were further inspected by statistical group comparison. That is, the two groups were statistically compared on those variables to see whether the groups systematically differed from each other in the errors made. The distribution of those variables were visually inspected using histograms, which turned out to be far from normal for most variables in both groups. Due to the non-normality of the majority of the variables, Kruskal-Wallis tests were conducted testing the distributions of both groups on the variables of interest. Due to the multitude of comparisons that resulted, Bonferroni-correction of the *p*-value was not applied, as the resulting *p*-value would be too conservative and a very conservative *p*-value lowers the statistical power, especially since power was diminished in the current study based on the small sample size and the expected small effect.

Results

Questionnaires.

The mean AQ score in the ASD group ($M = 28.58$, $SD = 10.24$, Min. = 10, Max. = 46) was significantly higher than of the controls ($M = 17.33$, $SD = 4.33$, Min. = 11, Max. = 26) as indicated by an independent-samples *t*-test, $t(14.82) = -3.51$, $p = .003$. However, not all participants in the ASD group scored above the recommended cut-off of 26 for a potential ASD (Woodbury-Smith et al., 2005). Since the AQ addresses only behaviour and interest and no other characteristic of ASD, it is possible that the items in the AQ did not address the symptomatology of all individuals and thereby biased the scores. Nevertheless, the parent ratings pointed towards an ASD in all participants. Given a cut-off of 15 and that the minimum score in the ASD group was 19 (Max = 30), it is indicated that all participants in the ASD group show autistic behaviour; although the parent data of two participants were missing (*n* = 10). Overall, with a mean of 22.60, the ASD group (*SD* = 4.01) scored above the cut-off showing they meet the criteria for an ASD diagnosis as reported by their parents.

Neutral recognition.

The mean accuracy of response in the control group for recognition of neutral faces was .90 (*Mdn* = .90, *SD* = .08) and the mean accuracy of response for neutral in the ASD group was .89 (*Mdn* = .90, *SD* = .11). Results of the Mann Whitney *U* test showed that the distribution of accuracy of response for neutral was the same across the ASD group and the controls (*U* = 67.50, *z* = -.26, *p* = .793), therefore, the groups did not differ in their labelling of neutral faces.

Affective state check.

The two groups showed a trend for differences in regards to their arousal rating at the start of the experiment (ASD: *Mdn* = 3, *ran* = 2; controls: *Mdn* = 2, *ran* = 2), *U* = 34, *z* = -1.67, *p* = .067. After the short documentary, the groups did not differ on their arousal ratings anymore (ASD: *Mdn* = 2, *ran* = 2; controls: *Mdn* = 2, *ran* = 3), *U* = 55.50, *z* = -1.02, *p* = .399. At the start of the experiment the two groups showed a trend for differences in valence (ASD: *Mdn* = 4, *ran* = 3; controls: *Mdn* = 4, *ran* = 2), *U* = 42.50, *z* = -1.92, *p* = .072. After watching the neutral documentary, however, there was no sign of a meaningful difference (ASD: *Mdn* = 3, *ran* = 3; controls: *Mdn* = 4, *ran* = 2), *U* = 61, *z* = -.71, *p* = .614. Both groups underwent the facial emotion recognition task with no difference in their affective state.

Accuracy of response.

The results of the GLMM showed a significant main effect of *group* (F(1,594) = 6.80, *p* = .009), with the controls (*M* = .76, *SE* = .02) outperforming those with ASD (*M* = .63, *SE* = .05); see Figure 20 .

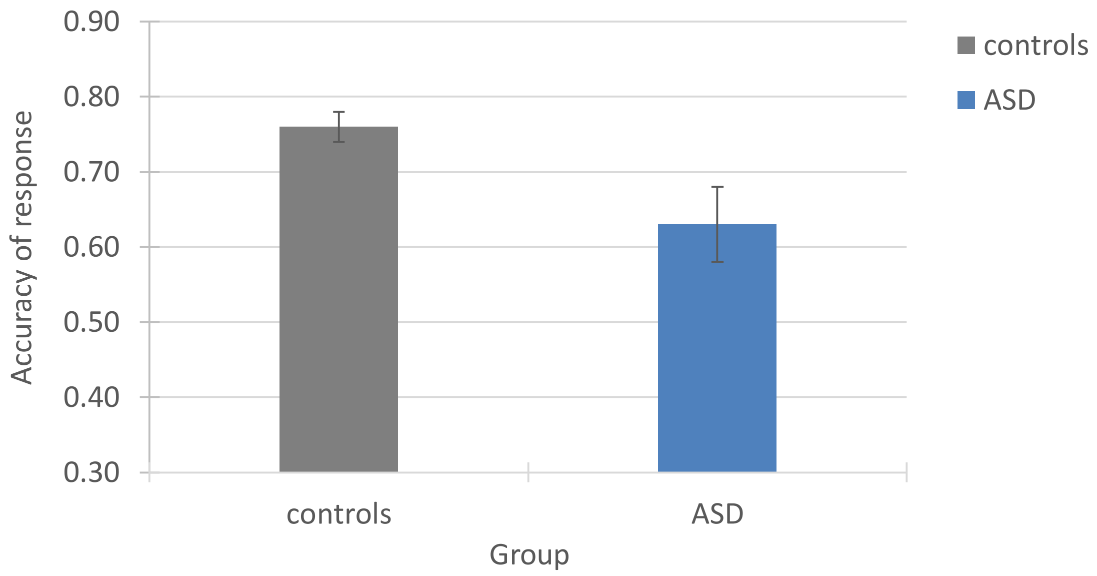


Figure 20. Overall accuracy of response by the ASD group and the controls. Accuracy of responses are expressed in decimal fractions. Error bars represent standard errors of the means.

The main effect of *emotion* was significant ($F(8,594) = 25.18$, $p < .001$). The main effect of *intensity* was also significant ($F(2,594) = 109.78$, $p < .001$). The interaction of *emotion***intensity* ($F(16,594) = 15.65$, $p < .001$) was also significant, i.e. the accuracy rates for the emotions were differently influenced by intensity of expression. Since these effects do not address the aims of the current study, the results of the pairwise comparisons, means, and SE are presented in Appendix E.

The interaction of *group***intensity* ($F(2,594) = 0.74$, $p = .479$) and the interaction of *group***emotions* were not significant ($F(8,594) = 1.23$, $p = .279$).

The interaction of *group***emotions***intensity* was significant ($F(16,594) = 1.94$, $p = .015$). Simple contrasts revealed that controls outperformed the ASD group in recognition of *anger* at low ($t(594) = 2.92$, $p = .004$) and intermediate ($t(594) = 3.44$, $p = .001$) intensity, but not at high intensity ($t(594) = 1.61$, $p = .108$). The control group achieved significantly higher accuracy rates on recognition of *fear* at low ($t(594) = 2.32$, $p = .021$), intermediate ($t(594) = 2.83$, $p = .005$), and high intensity ($t(594) = 3.26$, $p = .001$). The same pattern was found for *sadness* (low: $t(594) = 3.34$, $p = .001$; intermediate: $t(594) = 2.35$, $p = .019$; high: $t(594) = 2.40$, $p = .017$). *Embarrassment* was recognised better by the controls than ASDs only at low intensity ($t(594) = 2.13$, $p = .033$), no significant differences emerged at intermediate ($t(594)$

= 1.30, *p* = .193) or high intensity (*t*(594) = 1.05, *p* = .294). At no intensity level significant differences were found for *contempt* (low: *t*(594) = 1.04, *p* = .301; intermediate: *t*(594) = 1.21, *p* = .225; high: *t*(594) = .1.07, *p* = .285), *pride* (low: *t*(594) = 0.24, *p* = .813; intermediate: *t*(594) = .67, *p* = .097; high: *t*(594) = 1.66, *p* = .098), *disgust* (low: *t*(594) = 1.06, *p* = .290; intermediate: *t*(594) = 0.65, *p* = .518; high: *t*(594) = 0.77, *p* = .440), *happiness* (low: *t*(594) = -0.40, *p* = .692; intermediate: *t*(594) = 0.16, *p* = .874; high: *t*(594) = -0.14, *p* = .890), or *surprise* (low: *t*(594) = 0.54, *p* = .592; intermediate: *t*(594) = 0.70, *p* = .484; high: *t*(594) = -0.39, *p* = .697). In short, intensity of expression influenced the accuracy rates of the two groups differently for the emotions anger, fear, sadness, and embarrassment but not for contempt, pride, disgust, happiness, and surprise when investigated at each intensity level; see Figure 21 (means and standard errors are presented in Table 9).

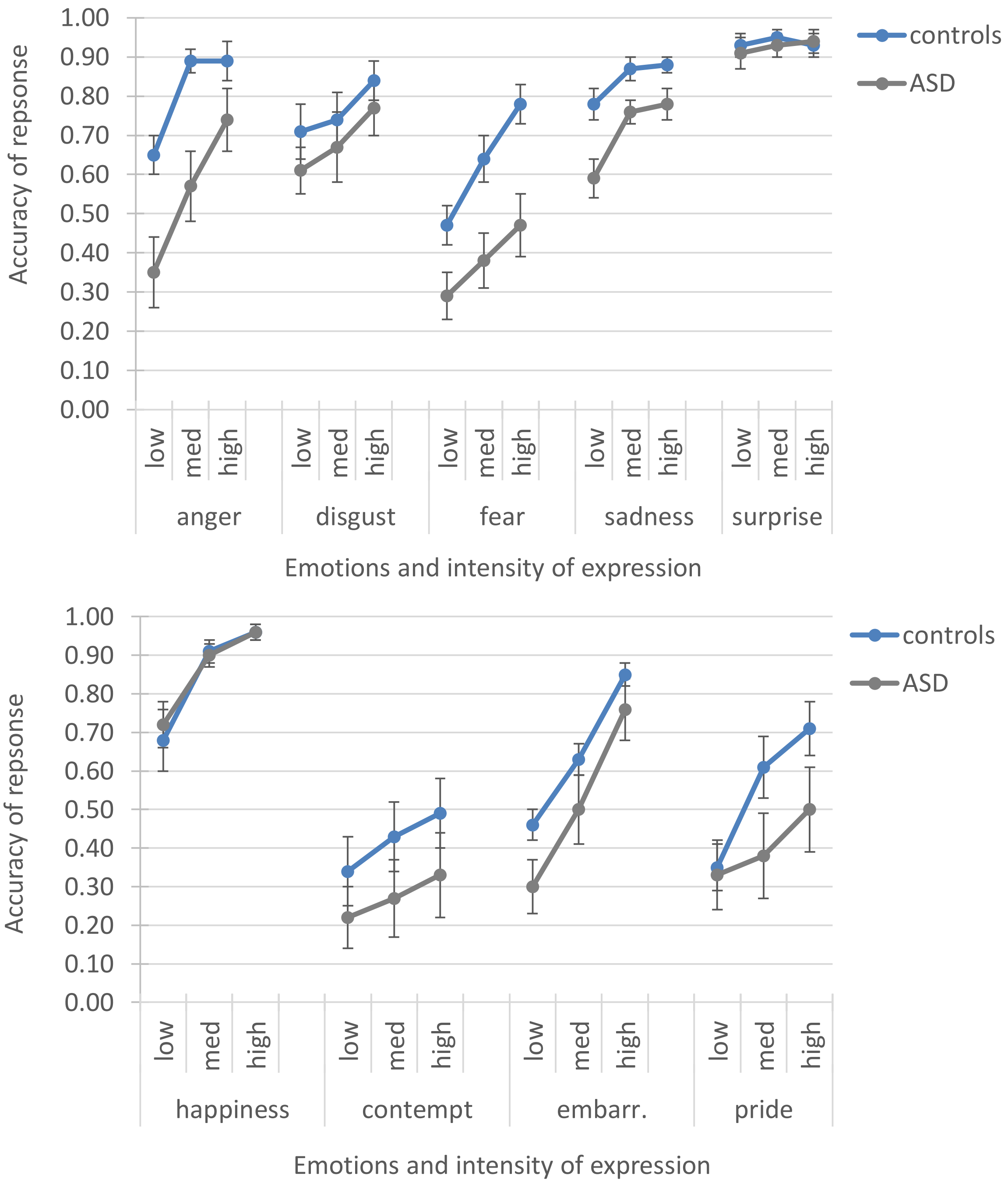


Figure 21. Accuracy of responses for the ASD group and the controls for the emotion categories of the ADFES-BIV at each intensity level. Accuracy of responses are expressed in decimal fractions. Error bars represent standard errors of the means.

Table 9

Accuracy of Response for Each Emotion at the Three Levels of Intensity for the ASD Group and Controls

	ASD (*n* = 12)						Controls (*n* = 12)					
	low		med		high		low		med		high	
Emotion	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
Anger	.35	.09	.57	.09	.74	.08	.65	.05	.89	.03	.89	.05
Disgust	.61	.06	.67	.09	.77	.07	.71	.07	.74	.07	.84	.05
Fear	.29	.06	.38	.07	.47	.08	.74	.05	.64	.06	.78	.05
Sadness	.59	.05	.76	.03	.78	.04	.78	.04	.87	.03	.88	.02
Surprise	.91	.04	.93	.03	.94	.03	.93	.03	.95	.02	.93	.03
Happiness	.72	.06	.90	.03	.96	.02	.68	.08	.91	.03	.96	.02
Contempt	.22	.08	.27	.10	.33	.11	.34	.09	.43	.09	.49	.09
Embarr.	.30	.07	.50	.09	.76	.08	.46	.04	.63	.04	.85	.03
Pride	.33	.09	.38	.11	.50	.11	.35	.06	.61	.08	.71	.07

Note. Means (*M*) and standard errors of the means (*SE*) are expressed in decimal fractions.

Confusion matrices and comparisons.

Visual inspection of the confusion matrices for the low intensity facial emotional expressions revealed that 12 confusions occurred more often than the 10% chance level of responding (see Table 10). Six of the confusions relate to sensitivity as they were between an emotion category and neutral. The other six confusions relate to specificity as they were between emotion categories. *Disgust* at low intensity was perceived as *neutral* by the ASD group in 15% of the presentations of disgust, whereas this confusion was made by the control group in 7% of the presentations of disgust and the results of the Kruskal-Wallis test showed that the distribution in both groups was indeed significantly different ($H(1) = 4.21$, $p = .040$). The confusion rate for *sadness* as *neutral* at low intensity was 24% in the ASD group and 6% in the controls and the results of the Kruskal-Wallis test showed that the distribution was significantly different in both groups ($H(1) = 7.28$, $p = .007$). *Anger* at low intensity was perceived as *neutral* by the ASD group in 28% of the anger presentations and in 13% by the controls; this difference in the distributions showed a trend ($H(1) = 3.45$, $p = .063$).

Contempt at low intensity was labelled as *neutral* in 45% of the contempt presentations by the ASD group and to 35% by the controls; the distribution was not significantly different ($H(1) = 0.53$, $p = .467$). The confusion rate of *embarrassment* at low intensity as *neutral* was 31% in the ASD group and 21% in the control group; the distribution was not significantly different ($H(1) = 0.88$, $p = .349$). The confusion rate of *happiness* at low intensity as *neutral* was 15% in the ASD group and 13% in the control group and the distribution was also not different ($H(1) = 0.25$, $p = .620$). In sum, the ASD group was more likely than the controls to perceive disgust, sadness, and anger as neutral when presented at low intensity. The ASD group perceived *anger* at low intensity as *disgust* in 14% of the anger presentations and the controls in 7% of the anger presentations, but the distribution in both groups was not significantly different ($H(1) = 1.50$, $p = .221$). The confusion rate of *contempt* at low intensity as *happiness* was 13% in the ASD group and 3% in the controls, but the distribution in both groups was comparable ($H(1) = 1.89$, $p = .170$). The confusion of *disgust* at low intensity with *anger* occurred to 20% in the ASD group and 19% in the controls; the distribution was not significantly different ($H(1) = 0.18$, $p = .673$). The confusion of embarrassment at low intensity with contempt was made in 13% of the *embarrassment* presentations in both groups and the distribution was not significantly different ($H(1) = 0.01$, $p = .928$). The confusion rate of *fear* at low intensity with *surprise* was 48% in the ASD group and 49% in the controls; the distribution of confusions was not significantly different ($H(1) = 0.09$, $p = .770$). The confusions rate of *pride* at low intensity as *happiness* was 56% in the ASD group and 53% in the control group; the distribution of confusions was not significantly different ($H(1) = 0.25$, $p = .620$). In sum, the groups did not differ in the confusions between emotion categories at low intensity.

Table 10

Confusion Matrices for the Low Intensity Facial Expressions for the ASD Group and Controls

ASD

	Emotions displayed (low intensity)										
Response	**Ang**	**Con**	**Dis**	**Emb**	**Fea**	**Hap**	**Neu**	**Pri**	**Sad**	**Sur**	**Total**
Ang	35	0	20	2	1	0	1	0	1	0	**25**
Con	8	24	3	13	0	6	5	5	3	0	**43**
Dis	14	0	60	3	5	1	0	0	5	0	**28**
Emb	1	4	0	31	7	1	1	0	1	1	**16**
Fea	3	1	0	2	30	0	1	0	4	3	**14**
Hap	0	13	0	10	0	71	1	56	1	1	**82**
Neu	28	45	15	31	9	15	89	6	24	6	**179**
Pri	0	10	0	1	0	7	1	33	1	1	**21**
Sad	8	1	1	5	1	0	1	0	58	0	**17**
Sur	3	3	0	1	48	1	0	0	2	90	**58**

Controls

	Emotions displayed (low intensity)										
Response	**Ang**	**Con**	**Dis**	**Emb**	**Fea**	**Hap**	**Neu**	**Pri**	**Sad**	**Sur**	**Total**
Ang	65	1	19	0	0	0	3	0	2	0	**25**
Con	8	35	3	13	1	6	1	3	1	0	**36**
Dis	7	1	70	2	1	0	0	0	4	0	**15**
Emb	0	8	0	47	1	1	1	1	3	0	**15**
Fea	0	0	0	3	47	1	1	0	5	5	**15**
Hap	0	3	0	3	0	68	0	53	1	1	**61**
Neu	13	35	7	21	1	13	92	6	6	1	**103**
Pri	1	7	0	1	0	9	0	35	0	0	**18**
Sad	6	9	0	10	0	1	3	0	77	0	**29**
Sur	1	1	1	1	49	1	0	1	1	93	**56**

Note. This table shows the correct responses (diagonal, in green) and confusions (above and below diagonal) in percentages for all emotions by group. Highlighted red are the confusions that occurred to a degree greater than chance. The last column shows the total of incorrect responses for each emotion across the 10 presented emotions. Ang = anger, Con = Contempt, Dis = Disgust, Emb = embarrassment, Fea = fear, Hap = happiness, Neu = neutral, Pri = pride, Sad = sadness, Sur = surprise.

Visual inspection of the confusion matrices for the high intensity facial emotional expressions revealed that 6 confusions occurred more often than the 10% chance level (see Table 11). Two of the confusions were between an emotion category and neutral and six between emotion categories. The confusion of *contempt* at high intensity with *neutral* occurred in 30% of the contempt presentations in the ASD group and 17% in the control group, but the distribution in both groups was not significantly different ($H(1) = 1.34$, $p = .248$). The confusion rate for *fear* at high intensity with *surprise* was 40% in the ASD group and 20% in the control group; the distribution was significantly different ($H(1) = 3.92$, $p = .048$). The confusion rate for *disgust* at high intensity as *anger* was 20% in the ASD group and 15% in the control group; the distribution was not significantly different ($H(1) = 0.29$, $p = .593$. The confusions rate for *pride* at high intensity as *happiness* was 48% in the ASD group and 27% in the control group, but the distribution was not significantly different ($H(1) = 2.66$, $p = .103$). The confusion of *contempt* at high intensity with *embarrassment* occurred in 10% of the contempt presentations in the ASD group and 13% in the control group, but the distribution in both groups was not significantly different ($H(1) = 0.84$, $p = .360$). The confusions rate for *embarrassment* at high intensity as *neutral* was 11% in the ASD group and 3% in the control group, but the distribution was not significantly different ($H(1) = 2.66$, $p = .103$). In sum, the ASD group confused fear with surprise significantly more often than controls at high expression intensity.

Table 11

Confusion Matrices for the High Intensity Facial Expressions for the ASD Group and Controls

ASD

	Emotions displayed (high intensity)										
Response	**Ang**	**Con**	**Dis**	**Emb**	**Fea**	**Hap**	**Neu**	**Pri**	**Sad**	**Sur**	**Total**
Ang	73	0	20	0	1	0	1	0	1	0	**23**
Con	10	34	2	4	0	0	1	1	0	0	**28**
Dis	4	1	76	1	8	1	0	0	3	1	**19**
Emb	3	10	0	74	3	0	1	0	3	1	**21**
Fea	1	1	0	0	47	0	2	0	3	3	**10**
Hap	0	9	0	6	0	96	1	48	1	1	**66**
Neu	3	30	1	11	1	1	91	1	8	0	**66**
Pri	0	10	0	0	0	2	0	49	1	0	**13**
Sad	4	1	1	3	1	0	1	0	76	0	**11**
Sur	1	3	0	0	40	1	1	1	3	94	**50**

Controls

	Emotions displayed (high intensity)										
Response	**Ang**	**Con**	**Dis**	**Emb**	**Fea**	**Hap**	**Neu**	**Pri**	**Sad**	**Sur**	**Total**
Ang	88	1	15	0	0	0	1	0	1	0	**18**
Con	5	49	1	6	0	0	3	2	0	0	**17**
Dis	3	1	83	0	2	0	1	0	4	0	**11**
Emb	0	13	0	85	0	1	0	0	1	0	**15**
Fea	2	0	1	2	78	0	1	0	3	7	**16**
Hap	0	3	0	1	0	96	0	27	0	0	**31**
Neu	0	17	0	3	0	1	89	1	3	1	**26**
Pri	0	2	0	1	0	3	0	70	0	0	**6**
Sad	2	10	0	3	0	0	5	0	88	0	**20**
Sur	0	3	0	0	20	0	0	0	0	92	**23**

Note. This table shows the correct responses (diagonal, in green) and confusions (above and below diagonal) in percentages for all emotions by group. Highlighted red are the confusions that occurred to a degree greater than chance. The last column shows the total of incorrect responses for each emotion across the 10 presented emotions. Ang = anger, Con = Contempt, Dis = Disgust, Emb = embarrassment, Fea = fear, Hap = happiness, Neu = neutral, Pri = pride, Sad = sadness, Sur = surprise.

Discussion

The current study set out to investigate facial emotion recognition in individuals with high-functioning ASD across a wider range of emotions and based on video stimuli including varying levels of intensity of expression. The current study identified an overall facial emotion recognition deficit in ASD compared to controls. The facial emotion recognition deficit in ASD was not influenced by expression intensity, as the deficits were consistent across the three levels of intensity. No group differences were found for specific emotions, although the recognition of the emotions included was influenced by intensity of expression to varying degrees between the ASD group and controls. Precisely, the ASD group showed deficits compared to the controls in recognition of anger at low and intermediate intensity, embarrassment at low intensity, and fear and sadness at all three intensity levels. Aim of the current study was also to qualify the deficit in facial emotion recognition in ASD by examining the confusions made and compare the two groups on these confusions. The ASD group showed greater impairments in specificity and sensitivity than the controls, but only for some emotions. Precisely, the ASD group showed deficits for fear recognition in regards to specificity next to deficits for recognition of anger, sadness, and disgust in regards to sensitivity. Overall, this study showed that there are deficits in facial emotion recognition in ASD when including varying levels of intensity of expression and further that diminished specificity as well as sensitivity are underlying. These confusions point towards an anomaly in face processing and impairments in visual perception in ASD.

Facial emotion recognition ability in ASD compared to controls.

By employing varying intensities of expression and a wider range of emotions significant group differences in facial emotion recognition between individuals with ASD and controls emerged in the current study, in line with the hypothesis. An overall deficit in facial emotion recognition in individuals with ASD when including varying intensities of expression is also in line with the published literature (Law Smith et al., 2010; Mazefsky & Oswald, 2007; Rump et al., 2009). That the deficit in facial emotion recognition in ASD was consistent across expression intensity level is not entirely in line with the reports by Law Smith et al. (2010) who found the deficit to be greatest at medium intensity, although the results from the current study align in so far that deficits show at each intensity level. It has to be noted that the reported interaction of group and intensity by Law Smith et al. (2010) was based on a trend and no clear significance. In contrast to the hypothesis and the results reported by Law Smith et al. (2010), the ASD group did not show significantly greater deficits in the recognition of specific emotions within the current study. The non-significant findings from the current study could be the result of the sample size, which was smaller than in the study by Law Smith et al. (2010). However, when considering the intensity of facial expression for each emotion, the accuracy rates achieved by the ASD group and the control group were differently influenced by intensity, which again is in line with the reports by Law Smith et al. (2010). It is possible that significant group differences for individual emotions only emerge when basic emotions are investigated, whereas the current study also included complex emotions. The findings show that there is an overall deficit in facial emotion recognition when including varying levels of intensity of expression and that emotion-specific deficits emerge depending on the intensity of expression. The latter is discussed in more detail in the following.

The current study found deficits in the ASD group for the recognition of the emotions anger, fear, sadness, and embarrassment that emerged at varying degrees of expression intensity. The result is in common with reports by Law Smith et al. (2010), as no significant group differences emerged for anger recognition at high intensity, but at the lower intensities. Unlike Law Smith et al. (2010), the current study did not yield significant group differences for the recognition of disgust and surprise at any intensity level. A possible explanation for the deviations in the results is that video recordings facilitate recognition compared to morphed sequences (which were used by Law Smith et al. (2010)). As discussed in previous chapters, alterations of the timings of an unfolding facial emotional expression, which naturally occurs using morphing, can affect perception based on the temporal characteristics that are embedded in our emotion representations (Bould et al., 2008; Kamachi et al., 2001; Sato & Yoshikawa, 2004). Disgust and surprise are both fast developing facial expressions and if the development of the facial expression is slower than typical for the emotion, then it is more difficult to recognise the emotion (see Recio et al., 2013). It is possible that for individuals with ASD this difficulty affects recognition rates more negatively than controls and could explain significant group differences based on morphed sequences as reported by Law Smith et al. (2010). It is further possible that video recordings offer temporal emotional information that is helpful for decoding of some emotions (e.g. surprise) to controls as much as to individuals with ASD.

That significant group differences were only found for negative emotions (anger, fear, sadness, and embarrassment - at certain levels of expression intensity) and not positive emotions leads to conclude that the valence (and intensity) of an emotion affects recognition performance by individuals with ASD. This result fits in with the literature, as valence and intensity have been identified in a literature review as core factors influencing recognition performance specifically in ASD (Nuske et al., 2013). However, the results from the current study are not in line with the hypothesis and reports that there is an emotion recognition deficit in ASD for *all* negative emotions (e.g. Ashwin et al., 2006), but in line with reports of impairments in the recognition of specific negative basic emotions (e.g. Wallace et al., 2008). There are less investigations on the recognition of complex emotions in ASD, although there are a few investigations on the understanding of embarrassment and deficits were reported (e.g. Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999; Capps, Yirmiya, & Sigman, 1992; Hillier & Allinson, 2002; Losh & Capps, 2006). Embarrassment is a social emotion that has the locus in how others perceive oneself and therefore requires an understanding of others' mental states. The latter has been reported to be deficient in ASD (Baron-Cohen, Leslie, & Frith, 1985; Heavey, Phillips, Baron-Cohen, & Rutter, 2000; A. M. Leslie & Frith, 1988; Reed & Peterson, 1990). It is possible that the lack of understanding of embarrassment extents to the recognition of its facial expression.

A potential explanation for deficits in the recognition of specific negative basic emotions is the distinctiveness of the facial emotional expressional. That is, some emotional facial expression are similar to each other due to featural overlap in their facial expression. For example, in fear and surprise the eyes are wide open and sometimes the mouth as well; the featural distinct aspect is the inner brow that is lowered in fear expressions but not surprise (Ekman & Friesen, 1978). The confusion of fear with surprise is a typical one, generally found in facial emotion recognition experiments due to the difficulty to process the subtle configuration in the brow (Roy-Charland, Perron, Beaudry, & Eady, 2014). To distinguish facial expressions of high featural overlap, holistic face processing is necessary and subtle configurations are necessary to process for correct recognition (e.g. Gauthier & Tarr, 1997; Mondloch, Le Grand, & Maurer, 2002). If the locus of focus is on a/the 'wrong' single feature (e.g. wide open eyes), then differentiation between emotions is diminished. It has indeed been reported that individuals with ASD rely on single feature processing more than controls (e.g. Behrmann et al., 2006; Doi et al., 2013), which also is in line with the postulation of the weak central coherence theory for individuals with ASD (Frith, 1989/2003; Happé & Frith, 2006) suggesting that individuals with ASD focus more on details than typical individuals. The lower recognition of featurally similar facial expressions (e.g. fear, as similar to surprise, and anger, as similar to disgust) in the ASD group within the current study might be the result from a focus on the most dominant yet misleading facial feature in those expressions, whereas the controls processed the face more holistically. Calvo and Nummenmaa (2008) identified configural face processing as the necessary strategy for recognition of fear and anger. The findings suggest that focus on individual features is less beneficial within facial emotion recognition than holistic and configural face processing and may contribute to the deficits shown in ASD. Combining this study with eye-tracking would give great insight into this assumption.

Specificity and sensitivity in facial emotion recognition.

Featural overlap in facial expressions can make recognition more difficult and therefore lead to lower accuracy rates. Confusions of facial emotional expressions are essentially underlying those lowered accuracy rates. This leads to the assumption that individuals with ASD are more prone to such confusions manifesting in significant group differences in accuracy rates. To investigate the assumption that individuals with ASD might differ from controls in their confusions made, both groups were compared on their confusions within the current study. In both groups confusions of featurally similar emotional facial expressions were found, as both groups confused fear with surprise, disgust with anger, and pride with happiness. The former two confusions are in line with previous reports (Humphreys et al., 2007; Jones et al., 2011). The confusion rates of disgust with anger and pride with happiness were not significantly different for the groups. However, the ASD group showed greater impairments of specificity in regards to recognition of fear. This is because individuals with ASD confused fear with surprise at high intensity significantly more often than controls even though emotional cues (features and configurations) become more apparent in higher intensities. This fits in with reports by Kennedy and Adolphs (2012) who found individuals with high-functioning ASD to judge emotional facial expressions less specific than controls, i.e. more overlapping with emotions other than the target emotion. Confusions due to diminished specificity are in line with the weak central coherence theory (Frith, 1989/2003); a focus on the most salient facial feature in ASD rather than holistic and configural face processing. It is further possible that underlying the decreased specificity are weaker prototypical mental representations of the emotional categories in ASD, although abnormal neurological processing has also been reported for example for fear (Ashwin, Baron-Cohen, Wheelwright, O'Riordan, & Bullmore, 2007). It seems like there is a general recognition deficit for fear in ASD due to diminished specificity that is independent of expression intensity.

Whereas a diminished specificity in facial emotion recognition suggests that emotional content was perceived yet misinterpreted, the perception of an emotional facial expression as neutral suggests a failure to perceive the emotional content. The ASD group stood out by confusing the facial expressions of disgust, sadness, and anger at low intensity with a neutral expression significantly more often than the controls. However, when the emotional cues were more intense (i.e. higher expression intensity), individuals with ASD did not differ from controls in number of those confusions. Wallace et al. (2011) found diminished sensitivity in ASD compared to controls over the six basic emotions combined and on emotion-level for sadness; in line with the confusions from the present study. For the emotions that were perceived as neutral by the ASD group at the lower intensities it can be concluded that the sensitivity towards those emotional cues is diminished compared to controls. An explanation for an impairment in facial emotion recognition is that individuals with ASD have a perceptual deficit. Dakin and Frith (2005) have accumulated evidence in a literature review that individuals with ASD are impaired in regards to motion perception. In line with that, Robertson et al. (2014) showed that in individuals with ASD performance in motion perception decreased more when viewing times were short compared to controls. Since facial expressions are dynamic and fleeting, impaired motion perception could manifest in confusions of emotional facial expressions with neutral ones. It is possible that these characteristics of facial expressions are what makes individuals with ASD less sensitive to detection of emotional content in the face, especially at the low intensity level where movements are rather small.

The results from the current study suggest that there is not one specific cause for a facial emotion recognition deficit, rather, anomalies in several processes of perception and face processing manifest in an emotion recognition deficit in ASD. The results from the current study further suggest that the anomalies in those processes in ASD influence the recognition of individual emotions differently and are additionally affected by factors such as expression intensity and valence of the emotion. Given facial emotion recognition is defined by firstly being able to detect the presence of an emotional expression in the face (sensitivity) and secondly being able to correctly identify the emotional information (specificity), the current study showed that individuals with high-functioning ASD display impairments in both aspects. It was demonstrated that individuals with ASD make typical confusions of featural overlapping emotions, but to an increased degree for fear, and have decreased sensitivity to the detection of the emotions sadness, anger, and disgust from faces. This information has implication for social trainings where facial emotion recognition is practised and should be adapted to the different needs of sensitivity and specificity.

Limitations.

There are limitations to the current study. The AQ was applied as the sole self-report measure to assess autistic traits. For some individuals from the ASD group the AQ failed to reliably assess autistic traits indicated by sum scores below the suggested cut-off, whereas the parent reports based on the SCQ was indicative of autistic traits in all individuals. A potential explanation is that the AQ focusses on the stereotypical image of an individual with high-functioning ASD. For example a strong interest in numbers and technology. However, interests and the behaviours of individuals with ASD can vary greatly, for example strong interest in arts. Further, some of the characteristics of ASD, like repetitive behaviour or sensual hypersensitivity, are simply not touched upon within the AQ. Variations in the phenotype of ASD can explain why some participants with ASD scored below the cut-off for ASD on the AQ. The gold standard in confirming an ASD diagnosis is the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012). However, ADOS assessment is costly and time consuming. The ADOS takes 40-60 minutes to administer and has to be carried out by a certified examiner. Due to the lack of certification (a very intensive and expensive procedure), a lack of financial resources to outsource ADOS testing, and the restricted time available for testing, the ASD participants within this research project were not ADOS tested. A short screening instrument applicable for research purposes (not diagnostics) based on self-report including items referring to all main characteristics of ASD is needed. Since all participants were diagnosed with an ASD by a clinical professional at some stage in their lives and the parent reports confirmed a current ASD diagnosis for all participants, it can be assumed that the participants really were from the ASD population.

Ideally, an IQ assessment would be been carried out with all participants. This is of relevance as IQ has been found to predict facial emotion recognition in ASD (Buitelaar, Wees, Swaab-Barneveld, & Gaag, 1999). Unfortunately, IQ assessment was unfeasible for the current research project, as research participation was only a minor aspect of the Autism Summer School and hence time for research was restraint. However, by recruiting the control group from the University of Bath Open Day and high-functioning ASD participants, it was assumed that all participants would be comparable on their IQ.

The study presented here had a sample size of 12 individuals with an ASD, which included 3 females. The unbalanced sex ratio reflects the sex ratio in the ASD population, which is 1 in 4 (Fombonne, 2005). The small sample size is a reflection of the difficulty to recruit individuals with ASD. As a result of the small sample size, the study presented here was underpowered. This also means, had a Bonferroni-correction been applied for the comparisons of the confusions, none of the differences had reached significance. The problem with Bonferroni-correction is that it is too conservative when a higher number of comparisons is conducted and that it disregards the literature and underlying theory. The results from the current study are generally in line with the autism literature. The results from the current study contribute to the literature by offering differentiated insights into facial emotion recognition including confusions. Hopefully the current study will spark further research using video-based stimuli of varying intensities to replicate the current findings on a larger sample or to further the results. Future research should also seek to combine video-based facial emotion recognition with eye-tracking and/or brain imaging to investigate the location of the diminished sensitivity and specificity of facial emotion recognition in ASD.

So far, facial emotion recognition has been investigated on a behavioural level within the current research project and it was of main interest to examine the influence of expression intensity on facial emotion recognition. Since throughout the studies presented here, people achieved lower recognition rates to low intensity facial emotional expressions than high intensity expressions, this leads to the question why. Based on the theoretical background given in Chapter 2 on the simulation theory, especially the reverse simulation model, it could be assumed that less facial mimicry occurs in response to low intensity facial expressions than high ones and that this reduced proprioceptive feedback is what drives the lower recognition rates. The facial emotion recognition task was hence combined with physiological measures of face EMG. The study investigates the role of proprioceptive feedback from face muscles within facial emotion recognition to find out more about the mechanisms involved in facial emotion recognition and is presented in the following chapter.

CHAPTER 9—Facial mimicry and the effect of proprioceptive feedback on facial emotion recognition

Introduction

The occurrence of facial mimicry is well documented, which involves subtle imitation of the emotional expressions on a subliminal level (e.g. Achaibou et al., 2008; Dimberg & Thunberg, 1998; Lundqvist & Dimberg, 1995; see Chapter 2). Nonetheless, the published research of mimicry of facial expressions has usually only included two or three face muscle sites for EMG assessment in each investigation. It is difficult to draw conclusions about facial mimicry in the whole face when only a limited number of face muscles are assessed, especially since the two muscles commonly included are the corrugator and the zygomaticus (review by Hess & Fischer, 2013). Those two face muscles can be involved in any display of negative or positive emotion respectively. Therefore facial mimicry is not necessarily emotion-specific and might be valence-based. Another problem with activity in the corrugator is that it also reflects cognitive load not associated with facial mimicry (Hess, Philippot, & Blairy, 1998). As a consequence, the literature on facial mimicry of specific facial features in response to emotional facial expressions is limited.

The reverse simulation model assumes that observing facial emotional expressions leads to facial mimicry. Since facial mimicry is assumed to be based on simulation processes of observed facial movement due to mirror neuron activity (see Chapter 2), there is reason to believe that the observed intensity of facial feature activation reflects relatively in the resulting intensity of facial mimicry. The reverse simulation model assumes further that facial mimicry elicits the observed emotion in the observer based on proprioceptive feedback from the face muscles and the generated information is then used to help identify the observed emotion. This suggests that there is a relationship between facial mimicry and facial emotion recognition. Previous chapters of this dissertation showed that the accuracy of response rates are lower for recognition of the lower intensity facial expressions than for the higher intensity facial expressions. Combining the assumption from the simulation model with the pattern of accuracy of response from varying intensities of facial expression, a possible explanation for the lower recognition rates of the lower intensity expressions is that less facial mimicry occurs in response to low intensity facial expressions. That is because low intensity facial mimicry results in less proprioceptive feedback from the face that could be utilised for decoding emotional expressions. However, nothing is known about the intensity of facial mimicry in relation to the intensity of observed facial emotional expressions.

The literature investigating the link between facial mimicry and facial emotion recognition is mainly behavioural, but seems to support a link as suggested by the reverse simulation model (see also Chapter 2). For example, Oberman, Winkielman, and Ramachandran (2007) conducted a study on facial emotion recognition and manipulated facial movement. The emotions investigated were happiness, disgust, fear, and sadness. A within-subject design was applied where in one experimental condition, no face movement manipulation was conducted, and in another facial mimicry was 'blocked' by having participants bite on a pen without the lips touching it. The authors showed in another experiment using face EMG that this face manipulation induces muscle activity that is incompatible with facial mimicry in response to some of the presented facial emotional expressions. As a result of the face movement manipulation, Oberman et al. (2007) found recognition of the emotions happiness and disgust to be diminished compared to the no face movement manipulation condition. Oberman et al. (2007) showed that recognition of emotion can be selectively impaired by blocking facial mimicry. The implication from this finding is that facial mimicry is a component of facial emotion recognition. However, it is also possible that biting on a pen and the resulting muscle activity interfered with recognition based on the proprioceptive feedback that was incongruent with the expressions of the presented emotions. It would be informative to investigate facial mimicry and the effects of the face movement manipulations on a physiological level in relation to facial emotion recognition.

However, further evidence for selective impairment of facial emotion recognition as a result of face movement manipulations comes from a study conducted by Ponari, Conson, D'Amico, Grossi, and Trojano (2012). They conducted two experiments where participants' face movements were manipulated and the effects on recognition of the six basic emotions were investigated. In one experiment, either the muscles of the upper half or the lower half of the face involved in emotional expressions were manipulated, next to a no face manipulation condition. One group bit on a chopstick horizontally (without the lips touching it) in order to block any facial mimicry in the lower part of the face. The other group had two little stickers attached at the inner edge of the eyebrows and were instructed to push them together in order to block any facial mimicry in the upper part of the face. Even though the word 'blocking' is used, the manipulations chosen were supposed to create constant muscular activity, which is not associated with any specific facial emotional expression, rather than a complete absence of muscular activity. The muscular activity was intended to interfere with facial mimicry and hence recognition. The group without facial movement manipulation was found to achieve higher accuracy rates than the groups with blocked facial mimicry. This result shows that interfering with facial mimicry diminishes recognition. A closer inspection of the influence of the face movement manipulations on each emotion showed that the face manipulations interfered with facial mimicry of the facial features that are considered most salient for specific emotions (i.e. the most noticeable and important for recognition) and decreased the accuracy rates for the respective emotions. Precisely, accuracy rates for happiness and disgust were lower in the lower face movement manipulation than in the other two experimental conditions. This finding is in line with the reports from Oberman et al. (2007) based on their face movement manipulation and means that biting on the chopstick particularly interfered with recognition of happiness and disgust, probably due to prevention of facial mimicry in the mouth region. This is interesting, as for both emotions the salient facial feature of the facial expression is situated in the lower part of the face (mouth and nose respectively) (Calvo & Nummenmaa, 2008; Khan, Meyer, Konik, & Bouakaz, 2012; Leppänen & Hietanen, 2007). The accuracy rates for recognising anger were lower in the upper face movement manipulation compared to the other two experimental conditions. The upper face movement manipulation seemed to hinder facial mimicry in the upper face and since anger is more salient in the upper part of the face (Calvo & Nummenmaa, 2008), recognition of anger was diminished. Recognition of fear was affected by both face movement manipulations compared to the no face movement manipulation, probably because the whole face has been found to be of salience for fear recognition (Calvo & Nummenmaa, 2008; Khan et al., 2012). These results are in line with the reverse simulation model in that facial mimicry (at least of the most salient facial feature) is needed for facial emotion recognition. The recognition of sadness and surprise in the study by Ponari et al. (2012) remained unaffected by the face movement manipulations. This result for surprise is unexpected, since the mouth has been identified as the most salient feature for its facial expression (Calvo & Nummenmaa, 2008; Khan et al., 2012); diminished accuracy rates should have therefore emerged in the lower face manipulation. That the recognition of sadness was unaffected by the face movement manipulations can be explained by the specific face muscle activations from both manipulations. The upper face with the two stickers as well as holding a chopstick with the teeth without the lips touching it might have resembled a facial expression of sadness. In this case, the proprioceptive feedback was in synchronisation with the activation resulting from making a sad face and did not block facial mimicry, which is why the recognition of sadness was not diminished by the face movement manipulations. The study by Ponari et al. (2012) showed that specific activations of facial muscles interfering with facial mimicry selectively impair facial emotion recognition. An alternative explanation would be that the facial movement manipulations distracted participants and diminished recognition. However, this is less likely, since the impairments were emotion-specific within each face movement manipulation. Therefore, the study suggests that facial mimicry is a necessary component of facial emotion recognition or at least that muscle activations that are incompatible with the observed facial emotion hamper recognition. Additional assessment of face EMG could have provided useful information on the effects of the face movement manipulations on facial mimicry.

The current study investigated facial mimicry using the ADFES-BIV. Employing short video stimuli is an improvement on previous research, which has used either static images or slow-moving morphed sequences. EMG measures of five different muscle sites in the face were taken in the current study to investigate if facial mimicry is emotion-specific or based on valence. The face muscle sites were: corrugator supercilii, zygomaticus major, levator labii, depressor anguli oris, and lateral frontalis (more details in the method section). This again is an improvement on previous research where generally two or three muscle sites have been investigated. With the ADFES-BIV the number of emotions investigated was increased to include the six basic and three complex emotions. Not much is known in regards to facial mimicry of complex emotions. By employing the ADFES-BIV, the intensity of facial emotional expression also varied, allowing investigation of the intensity of facial mimicry in relation to observed intensity of facial expression. To date, it is not known whether the intensity of facial mimicry varies in relation to the intensity of the observed facial expressions. To investigate the effects of proprioceptive feedback from the face on facial emotion recognition, a within-subject approach was taken applying three different experimental conditions: no face movement manipulation, explicit imitation of the observed facial emotional expressions, and a blocked facial mimicry condition. Accuracy of response was included as the DV.

Aims and hypotheses

The aims of the present study were to:

(1) test if a distinct pattern of EMG activation across the five muscle sites during facial mimicry and explicit imitation occurs for each of the six basic and three complex emotions by comparing the EMG responses of the five muscle sites for each emotion.

(2) investigate whether the intensity of observed facial emotional expressions reflects in the intensity of facial mimicry and explicit imitation in participants by comparing the EMG responses to the three intensity levels of the ADFES-BIV and neutral faces.

(3) examine if, as suggested by the reverse simulation model, proprioceptive face muscle feedback (especially from facial mimicry) is a means to facial emotion recognition. For this purpose, the recognition rates from three experimental conditions (from facial mimicry, explicit imitation, and blocked facial movements) were compared for each emotion included in the ADFES-BIV.

Hypotheses to aim 1:

(1) Based on facial features as specified in van der Schalk et al. (2011) for the ADFES based on the FACS (Ekman & Friesen, 1978; new edition: Ekman et al., 2002), the identified salience for specific emotions (Calvo & Nummenmaa, 2008; Khan et al., 2012) and previous reports on facial mimicry (Lundquist, 1995), it was expected that a specific pattern of face muscle activation would emerge for each emotion investigated.

Precisely:

(1.1) For the emotional facial expression of *happiness*, which includes a smile, an increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.2) For the emotional facial expression of *pride*, which includes a smile, an increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.3) For the emotional facial expression of *contempt*, which includes a unilateral smile and a unilateral pulled up eyebrow, an increase in zygomaticus and frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles (but only if participants activated the left side of the face).

(1.4) For the emotional facial expression of *anger*, which includes a frown, an increase in corrugator activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.5) For the emotional facial expression of *fear*, which includes a frown and pulling up the eyebrows, an increase in corrugator activity and lateral frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.6) For the emotional facial expression of *sadness*, which includes a frown and pulling down the mouth corners, an increase in corrugator activity and depressor anguli oris activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.7) For the emotional facial expression of *surprise*, which includes wide open eyes, an increase in lateral frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.8) For the emotional facial expression of *embarrassment*, which includes narrowing the lips, an increase in depressor anguli oris activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

(1.9) For the emotional facial expression of *disgust*, which includes a wrinkled nose, an increase in levator labii activity compared to baseline was expected that would be higher than the increase in activity in the other muscles.

Hypothesis to aim 2:

(2) It was expected that greater intensity of observed facial expressions would be evident with higher EMG responses during explicit imitating facial expressions. If the intensity of facial mimicry is also relatively influenced by the intensity of observed facial expressions, then greater intensity of observed facial expression would be evident in higher intensity of facial mimicry based on EMG responses.

Hypotheses to aim 3:

(3) It was expected that the proprioceptive feedback resulting from facial mimicry would be used during decoding of facial emotional expressions shown by greater accuracy of response when face movement was not restricted compared to the blocked facial mimicry experimental condition. It was expected that enhanced proprioceptive feedback during explicit imitation of facial expressions would facilitate recognition compared to the blocked facial mimicry experimental condition and the no facial movement manipulation condition. The effects of the face movement restriction were expected to be most prominent for the emotions with saliency on the mouth region (happiness, disgust); since the facial muscles of the upper face would be unaffected by the pen in mouth manipulation).

Method

Participants.

The sample consisted of 86 participants (43 male, 43 female), aged between 18 and 48 (*M* = 19.6, *SD* = 3.6). Due to technical equipment failure, the data of two participants was lost (1 male, 1 female), leading to a total *N* = 84. These consisted of 82 Undergraduate and 2 Postgraduate students of the University of Bath. The majority were from the Department of Psychology (*n* = 48), but also Engineering (*n* = 17), Computer Sciences (*n* = 4), Sciences (*n* = 4), Social Sciences (*n* = 3), Maths (*n* = 6), Economics (*n* = 1), and Management (*n* = 1). The study resulted in an ethnically diverse sample: British (*n* = 59), Chinese (*n* = 8), Russian (*n* = 2), Malaysian (*n* = 2), Romanian (*n* = 1), Mexican (*n* = 1), Belgian (*n* = 1), Ukrainian (*n* = 1), Swiss (*n* = 1), Taiwanese (*n* = 1), Egyptian (*n* = 1), South Korean (*n* = 1), French (*n* = 1), Italian (*n* = 1), Japanese (*n* = 1), Norwegian (*n* = 1), Hungarian (*n* = 1). However, all participants were assumed to be fluent in English, based on it being one of the entry requirements to study at the University of Bath. All participants had normal or corrected-to-normal vision to assure that the visual emotion cues could be perceived. Three of the participants reported a current clinical diagnosis of a mental disorder (1 depression and anxiety, 1 depression, 1 anxiety) and being on treatment medication. They were not removed since their recognition rates were comparable to the rest of the sample. Ethical approval for this study was given by the Department of Psychology Ethics Committee of the University of Bath (Ref: 14-008).

Stimuli.

The validated ADFES-BIV videos (Chapter 6) including all 10 emotions and the three intensity levels of expression were used as stimuli.

EMG recording.

The BIOPAC MP150 System with the Acqknowledge software (Version 4, Biopac Systems, Inc., Goleta, CA) was used for recording of EMG data. EMG was recorded with the EMG110C unit for each of the five facial muscle sites (corrugator supercilii, zygomaticus major, levator labii, depressor anguli oris, and lateral frontalis). Pairs of shielded surface silver-silver chloride (*Ag-AgCl*) electrodes (EL254S) with a contact area of 4mm diameter, filled with conductive gel (saline based Signa Gel), were used for data assessment. The EMG signal was amplified by 2000 and online bandpass filtering of 10Hz and 500Hz was applied to account for motion artefacts like eye blinks (Boxtel, 2001). Grounding was achieved through the VIN- of the TSD203. The sampling rate was 1000Hz throughout the experiment.

Facial action units, emotions, and associated muscles.

EMG activity of five face muscle sites were recorded. The facial action units for the emotions included in the ADFFES-BIV and associated face muscle sites investigated are presented in Table 12.

Table 12

Facial Action Units and Associated Face Muscle Sites for the ADFES-BIV Emotion Categories

Emotion	**AU**	**Description**	**Face muscle site**
anger	AU4	eyebrow lowered	corrugator
	AU5	upper lid raised	
	AU7	eyelids tightened	
	AU23	lips tightened	
disgust	AU9	nose wrinkled	levator
	AU10	upper lip raised	
	AU25	lips parted	
sadness	AU1	inner eyebrow raised	
	AU4	eyebrow lowered	corrugator
	AU15	lip corners depressed	depressor
fear	AU1	inner eyebrow raised	
	AU2	outer eyebrow raised	frontalis
	AU4	eyebrow lowered	corrugator
	AU5	upper lid raised	
	AU20	lips stretched	
surprise	AU1	inner eyebrow raised	
	AU2	outer eyebrow raised	frontalis
	AU5	upper lid raised	
	AU26	jaw dropped	
happiness	AU6	cheeks raised	
	AU12	lip corners pulled up	zygomaticus
contempt	AU1	inner eyebrow raised, unilateral	
	AU2	outer eyebrow raised unilateral	frontalis
	AU14	dimple, unilateral	zygomaticus
embarrassment	AU12	lip corners pulled up	zygomaticus
	AU14	dimple	
	AU23	lips tightened	depressor
pride	AU6	cheeks raised	
	AU12	lip corners pulled up	zygomaticus

Design.

A within-subject design was applied. The experiment consisted of four experimental conditions with differing face movement manipulations. (1) A passive viewing with instructions simply to passively watch the videos. (2) Watching the videos with an emotion recognition task for each video. (3) Watching the videos with an emotion recognition task and a task to block facial mimicry. (4) Watching the videos with an emotion recognition tasks along with instructions to explicitly imitate the presented emotion as seen. It was likely that accuracy rates would increase over the course of the experiment. These increases would not be attributable to the experimental manipulations but to non-specific improvement effects. This effect has been shown based on the data from the study in Chapter 6; see Appendix F. Therefore, the experimental conditions were balanced using an 'incomplete repeated measures design' in order to counter the non-specific improvement effects. As a result, 6 versions of the experiment were created. The passive viewing condition appeared first in all versions of the experiment, which was done to keep participants unaware for this part of the experiment of the study aims (that facial movements and facial emotion recognition were assessed). The remaining 3 experimental conditions were randomised across the versions of the experiment and participants were assigned to each version in a randomised order. The presentation of trials within the experiment was also randomised for each participant. E-Prime 2.0 was used as stimulus presentation software and to record accuracy of response.

Procedure.

The testing sessions took place in a laboratory at the University of Bath with only the experimenter present throughout the session. Participants gave informed consent prior to participation. Before EMG electrode attachment, the relevant parts of the face were cleaned with alcohol swabs to remove make-up, oil of the skin, and dead skin and assure the best possible electrode contact with the skin. The 10 face EMG electrodes were then placed in pairs over the respective muscles according to the guidelines by Fridlund and Cacioppo (1986); see Figure 22. In accordance with these guidelines, the electrodes were all placed on the left side of the face, based on the fact that humans tend to express emotions more intensely with the left side (Sackeim, Gur, & Saucy, 1978). Each pair of electrodes was placed in close proximity (1cm between the electrode centres) with double-stick adhesive rings. The lower electrode on the lateral frontalis was placed in line with the pupil of the eye when looking straight ahead about 1cm above the eyebrow and the second electrode was placed just above. For the corrugator supercilii one electrode was placed directly above the eyebrow lined up with the inner commissure of the eye fissure and the other electrode was placed lateral and slightly above the first one above the eyebrow. For assessment of the levator labii one electrode was placed lateral to the baseline of the nostril and one electrode right below and slightly lateral of the first one. To assess zygomaticus major activity, one electrode was placed in between the corner of the mouth and the bony dimple next to the ear and one electrode toward the mouth on the same imaginary line. The first electrode on the depressor anguli oris was placed about 1cm below the corner of the mouth and the second below the first one. Participants were told that measures of temperature and sweat response would be taken with the electrodes.

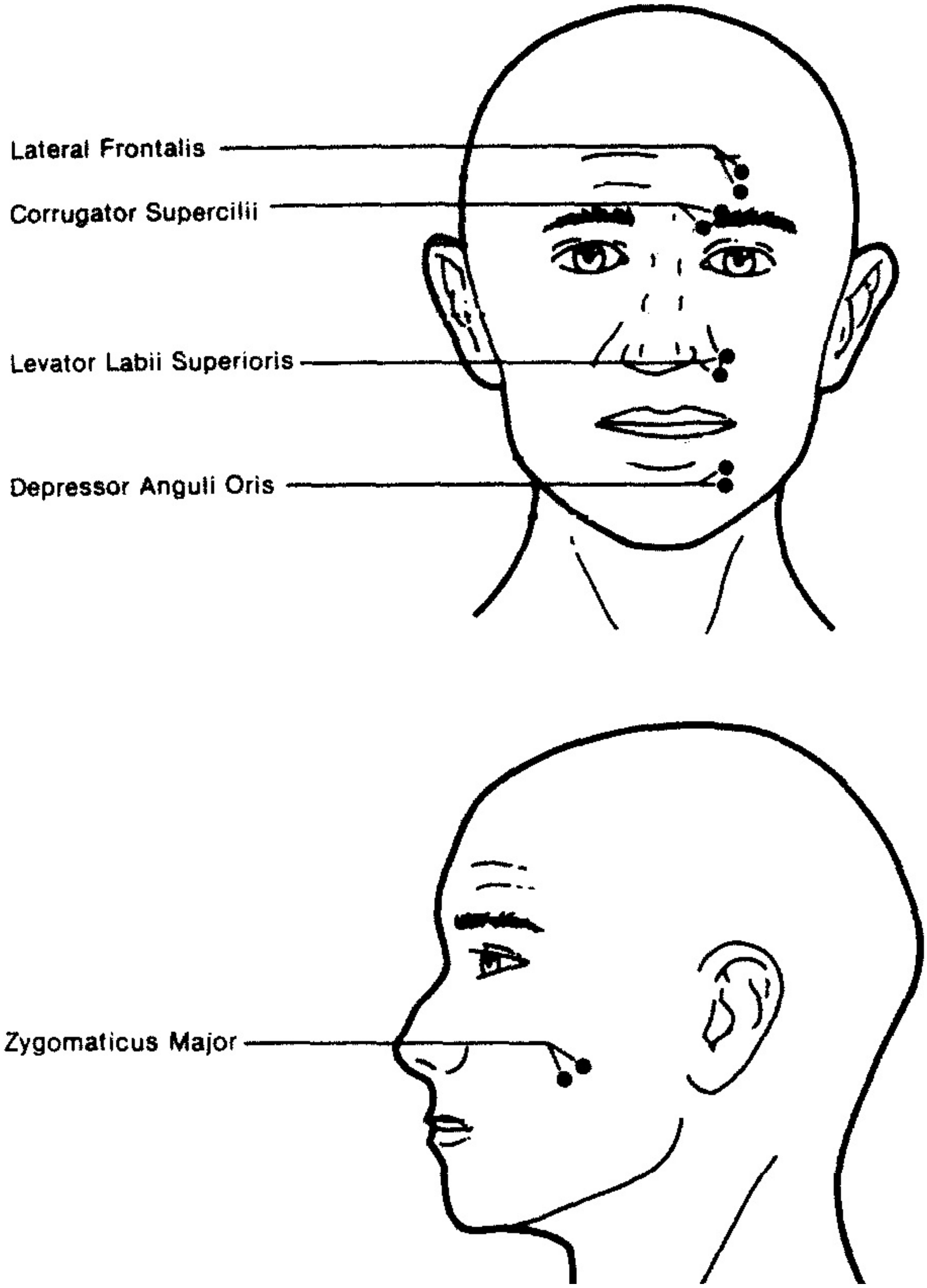


Figure 22. Electrode placement for assessment of face EMG according to Fridlund and Cacioppo (1986). The pictures show the electrode placement for the face muscles investigated in the current study.

Once all electrodes were attached, the experiment was started. Participants were seated approximately 60cm from the screen at eye level. A 17'' monitor was used with a screen resolution of 1024x768 pixel, matching the resolution of the ADFES-BIV videos. Just as in the studies presented in previous chapters, the experiment started with taking an affective rating before and after the neutral video clip, after which the emotion processing part of the experiment started. The experimenter instructed the face movement manipulations verbally before each of the four experimental conditions of 90 trials (360 videos/4 experimental conditions = 90 videos). For the no face movement manipulation condition without facial emotion recognition participants were asked to simply watch the videos. Afterwards, participants did 10 practice trials for the facial emotion recognition task and were instructed to choose the label with the emotion term out of the given ones they felt suited best the observed expression. For the no face movement manipulation condition combined with a facial emotion recognition task, the participants were instructed to watch the videos and choose a label just as during the practice trials. For the blocked mimicry manipulation, participants were instructed to hold a pen between their lips (not their teeth) with it sticking straight out of their mouth over the whole period of 90 trials. This method is commonly used (see Foroni & Semin, 2011; Oberman et al., 2007; Ponari et al., 2012). The exact pen holding position was adopted from Foroni and Semin (2011) with the intent to create muscular noise incompatible with the muscle responses involved in the emotional expressions displayed (Oberman, et al. 2007). The pen holding was demonstrated by the experimenter. Participants were further instructed to keep constant pressure on the pen. Participants were reminded to put pressure on the pen by the experimenter when the pen moved downwards. The block of trials was only started once the participant held the pen as instructed. Participants also had to choose an emotion label after each video. For the explicit imitation condition, participants were instructed to imitate the facial expressions exactly how they saw them simultaneous with the viewing, and choose a label afterwards. After completion of the experiment, the electrodes were taken off and participants filled out a short survey on-line via the Bristol Online Survey. As part of the survey, participants were asked if they had guessed the real purpose of the electrodes. Eight participants guessed correctly that explicit imitation might facilitate recognition. No one guessed the purpose of the pen, which was assumed to be a distractor rather than a means to block facial mimicry. Participants were then fully debriefed and were compensated with course credit or GBP 7.

Data reduction.

The Autonomic Nervous System Laboratory 2.6 (ANSLAB; Wilhelm & Peyk, 2005) was used for filtering the EMG data. The EMG signals were 50Hz notch filtered, 28Hz high-pass filtered, and the rectified signal was smoothed with a moving average width of 50ms. For each trial, the 200ms prior to stimulus onset was used as baseline (as often done in face EMG studies, e.g. Künecke, Hildebrandt, Recio, Sommer, & Wilhelm, 2014): means were extracted for this period using ANSLAB. The time from stimulus onset plus 2.1sec was used as the event windows: maximum values (peaks) extracted with ANSLAB for this period. With facial mimicry being discernible after approximately 400ms after facial emotional expression onset (Dimberg & Thunberg, 1998) and because it has been found to reach its maximum within 1 second (Achaibou et al., 2008), the event timing of 3.1s from stimulus onset allowed the capture of the peak reaction even if the emotional information became visible late in the video (e.g. for low intensity facial emotional expressions). EMG responses were calculated by subtracting the baseline mean from the event peak. Event peaks were taken and not event means, since the latter would not reflect well the EMG responses. This is because the event window was rather long and therefore included periods without a facial mimicry response, which would falsely lower the means.

Manipulation compliance check.

The EMG data were *z*-standardised within-subject for each face muscle site, but across experimental conditions to be able to compare the EMG responses between the experimental conditions on their original scale. Within-subject *z*-transformations reflect each individual's responses relative to their mean and standard deviation (Ben-Shakhar, 1985) and was necessary due to the individual differences that exist in EMG activity and reactivity (Bush, Hess, & Wolford, 1993; Cacioppo et al., 1992). Standardisation within muscle sites is recommended to account for the differences in EMG activity between muscle sites (Cram, 1998; Oberman et al., 2007). After *z*-transformations, the data were plotted for each participant and visually inspected for discernible EMG responses to confirm individual compliance with the face movement manipulations. All participants complied with the instructions and therefore no participant was eliminated. However, instead of the passive viewing (the experimental condition with facial emotion recognition), participant 26 did the explicit imitation condition twice, not following the instructions. Therefore the EMG data were removed for the latter experimental condition for this participant.

Data preparation an analysis.

Statistical analysis were conducted using IBM SPSS Statistics 22 (IBM Corp., Armonk, NY).

Face movement manipulations.

The raw EMG data were prepared in order to compare the EMG activity during the four experimental conditions to each other. The raw EMG responses for the individual trials were averaged across emotions to obtain variables for each of the five muscles per experimental condition. Those variables were then *z*-standardised across experimental conditions (within-subject and within-muscle) to examine whether or not differential EMG responses resulted from the face movement manipulations. The *z*-standardised data did not show any extreme values; all *z*-values fell within the range of -3 to +3 *SD*s from the mean and therefore no data was omitted. To examine whether or not the facial movement manipulations led to differing levels of EMG responses between manipulations, a 5 face muscle sites (corrugator, zygomaticus, levator, frontalis, depressor) x 4 experimental conditions (no face manipulation without facial emotion recognition, no face movement manipulation with facial emotion recognition, explicit imitation, blocked facial mimicry) repeated measures ANOVA was conducted. The assumption of a normal distribution was not met for 19 of the 20 variables according to Shapiro-Wilk tests (W's(83) < .96, p's < .05), but when sample sizes are equal the F-statistic is considered robust to this violation of normality (Field, 2009). This allowed comparison of the EMG activity of the four experimental conditions. Additionally, this allowed the identification of which muscles were activated by the blocking manipulation. The experimental conditions are abbreviated from herein as follows: no face manipulation without facial emotion recognition = SM (spontaneous facial mimicry), no face manipulation with facial emotion recognition = FER, explicit imitation with facial emotion recognition = IM, blocked facial mimicry with facial emotion recognition = BM.

Explicit imitation EMG pattern per emotion.

The raw EMG data was prepared for analysing emotion-specific EMG response patterns within a separate data file. The raw EMG responses of the individual trials from the IM condition were averaged to categories for each emotion and those categories *z*-standardised within experimental condition, subjects, and muscles. This allowed examination of the EMG pattern within the IM condition and accounted for inter-individual variability of EMG activity between participants and face muscle sites. The *z*-standardised data did not show any extreme values; all *z*-values fell within the range of -3 to +3 *SD*s from the mean and therefore no data was omitted. To examine the EMG patterns for the emotion categories, a repeated measures ANOVA was conducted for the experimental conditions of IM with the 10 emotions (anger, disgust, fear, sadness, surprise, happiness, neutral, pride, contempt, embarrassment) and 5 face muscle sites as within-subject factors. The assumption of a normal distribution was not met for 40 of the 50 variables according to Shapiro-Wilk tests (W's(84) < .96, p's < .05). However, transformations were not conducted since the data were already *z*-standardised and when sample sizes are equal the F statistic is considered robust to this violation of normality (Field, 2009).

Facial mimicry EMG pattern per emotion.

The raw EMG data was prepared for analysing emotion-specific EMG response patterns within a separate data file. The raw EMG responses of the individual trials from the SM condition were averaged to categories for each emotion and those categories *z*-standardised within experimental condition, subjects, and muscles. This allowed to examine the EMG pattern within the SM and accounting for inter-individual variability of EMG activity between participants and muscle sites. The *z*-standardised data did not show any extreme values; all *z*-values fell within the range of -3 to +3 *SD*s from the mean and therefore no data was omitted. To examine the EMG patterns for the emotion categories, a repeated measures ANOVA was conducted for the experimental conditions of SM with the 10 emotions and 5 face muscle sites as within-subject factors. The assumption of a normal distribution was not met for 42 of the 50 variables according to Shapiro-Wilk tests (W's(84) < .97, p's < .05). However, transformations were not conducted since the data were already z-standardised and when sample sizes are equal the F statistic is considered robust to this violation of normality (Field, 2009).

Intensity of observed expression and intensity of explicit imitation.

The raw EMG data was prepared for investigating whether the intensity of observed facial emotional expressions reflected relatively in the EMG responses during explicit imitation within a separate data file. The individual trials of the raw EMG responses from the IM condition were categorised into the three intensity levels and neutral within each of the five face muscle sites. Inspection of the categories with boxplots showed that in the IM condition there were two extreme values for the variable neutral in the corrugator (participants 3 and 69), two in the depressor during neutral faces (participants 73 and 78), one on frontalis during neutral faces (participant 78), two on levator during neutral faces (participants 7 and 70), and one on zygomaticus during neutral faces (participant 78). Shapiro-Wilk tests indicated a distribution significantly different from a normal distribution for all 20 variables with skews to the right (W's < .95, p's < .001). For seven of these variables normality could not be achieved by log transforming the data (W's < .97, p's < .05). Due to the robustness of the F statistic when sample sizes are equal (Field, 2009), a repeated measures ANOVA was conducted after the extreme values were adjusted. (The comparisons of the intensity levels were conducted at the level of intensity, because at the level of emotion by intensity only three trials existed, which did not allow for reliable building of categories).

Intensity of observed expression and intensity of facial mimicry.

The raw EMG data was prepared for investigating whether the intensity of observed facial emotional expressions reflected relatively in the EMG responses of facial mimicry within a separate data file. The individual trials of the raw EMG responses from the SM condition were categorised into the three intensity levels and neutral within each of the five face muscle sites. Due to the right-skew underlying the categories from the SM condition as identified by inspection of the histograms and the significant Shapiro-Wilk statistics (*W*'s < .91, *p*'s < .001) as well as to eliminate the multitude of extreme values identified via boxplots in the variables of the SM condition (muscle by intensity), all 20 variables were log transformed. After transformation there were no extreme values and 18 variables were not significantly different from a normal distribution according to Shapiro-Wilk tests (*W*'s > .97, *p*'s > .05). Examining skewness and kurtosis values only for zygomaticus at high intensity, there was a slight right-skew (skewness = .632, *SE* = .267, *z*-skewness = 2.37). All other *z*-kurtosis and *z*-skewness values were < 1.96. Therefore, a repeated measures ANOVA was conducted with the 5 face muscle sites and 4 intensities of expression (low, intermediate, high, and also neutral) as within-subject factors to examine whether or not the intensity of observed facial expressions of emotion reflected in facial mimicry. (The comparisons of the intensity levels were conducted at the level of intensity, because at emotion level only three trials existed by intensity, which did not allow for reliable building of categories).

Effects of proprioceptive feedback on facial emotion recognition.

The individual facial emotion recognition trials were combined to emotion categories within each of the three experimental conditions including a facial emotion recognition task resulting in 30 variables (10 emotions x 3 experimental conditions). None of the variables showed a normal distribution according to Shapiro-Wilk statistics (*W*'s(83) < .95, *p*'s < .005). Boxplots were inspected to identify extreme values. The identified extreme values were corrected to the value just above not consisting an extreme value (BM: neutral 2 extreme values, surprise 2 extreme values; FER: neutral 1 extreme value; IM: neutral 1 extreme value, surprise 2 extreme values). To keep the sample sizes exactly equal for each experimental condition, for subject 26 the data from the FER condition were replaced with the sample means for each emotion category. Due to the robustness of the *F* statistics when sample sizes are equal (Field, 2009), for each of the 10 emotions a repeated measures ANOVA with the three experimental conditions as within-subject factors was conducted.

Results

Face movement manipulation check.

Before the hypotheses could be tested, it was checked if the face movement manipulations worked. The results from the face muscle sites (5) x experimental conditions (4) repeated measures ANOVA on the data of the EMG responses indicated a violation of the assumption of sphericity for the main effect of experimental condition ($W(5) = .21$, $p < .001$) and the experimental condition*muscle interaction ($W(77) = .00$, $p < .001$), Greenhouse-Geisser adjustment of degrees of freedom was therefore applied. Due to the standardisation within face muscle sites, a main effect for the muscle sites could not be computed; all muscles had a mean of 0 and *SD* of 1. A significant *main effect of experimental condition* ($F(1.54,126.46) = 910.12$, $p < .001$, partial $\eta^2 = .917$, power = 1.00) was found. Pairwise comparisons showed that the EMG responses during the IM condition were significantly higher than during the BM, FM, and FER condition (p's < .001). The EMG responses during the BM condition were significantly lower than during the IM condition, but higher than during the SM and the FER condition (p's < .001). The EMG activity during the FER and the SM condition was not significantly different ($p = .632$). These findings reflect the desired pattern for manipulation of facial responses [IM > BM > FER = SM]; see Figure 23.

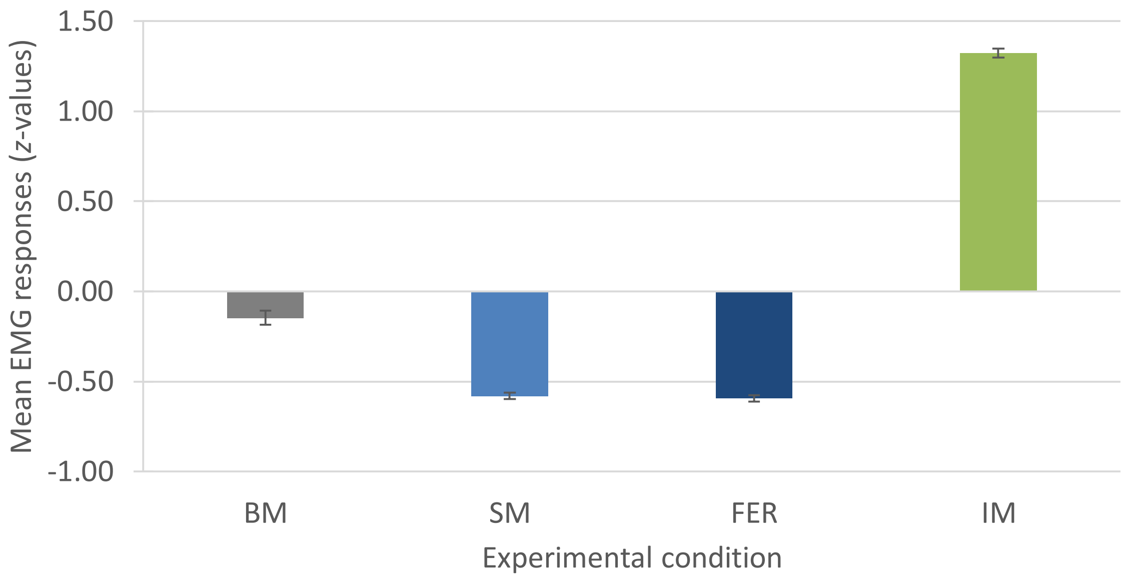


Figure 23. Mean EMG responses for each of the four experimental conditions. Since the data were *z*-standardised (across experimental conditions, within-subject, within-muscle), a mean of 0 does not reflect an absence of activity, but the average activity across experimental conditions. BM = blocked mimicry, FER = no face movement manipulation with facial emotion recognition, SM = no face movement manipulation (without facial emotion recognition), IM = explicit imitation. Error bars represent standard errors of the means.

The interaction of *face muscle site*experimental condition* was also significant ($F(5.33,436.91) = 29.87$, $p < .001$, partial $\eta^2 = .267$, power = 1.00); see Figure 24. Pairwise comparisons within each experimental condition between the face muscle sites showed that during the **IM** condition the activity in the corrugator was significantly different from the levator ($p = .035$) and the depressor ($p = .001$), but not frontalis ($p = .922$) or zygomaticus ($p = .734$). The activity in the frontalis was significantly higher than in the depressor ($p < .001$) and a trend was found in comparison to the levator ($p = .053$). The activity in the zygomaticus was significantly higher than in the depressor ($p < .001$) and the levator ($p = .017$). During imitation the activity in the levator was significantly higher than in the depressor ($p = .026$) [frontalis = corrugator = zygomaticus > levator > depressor]. During the **SM** condition the activity in the corrugator was significantly higher than in the depressor ($p < .001$) and levator ($p = .007$), but not frontalis ($p = .120$) or zygomaticus ($p = .311$). The activity in the depressor was significantly lower than in the frontalis ($p < .001$), zygomaticus ($p <. 001$), and levator ($p < .005$). The activity in the frontalis was significantly higher than in the levator ($p < .001$) and zygomaticus ($p = .031$). The activity in the zygomaticus was significantly higher than in the levator ($p = .019$) [frontalis = corrugator = zygomaticus > levator > depressor]. During the **FER** condition the EMG activity in the corrugator was significantly higher than in the depressor ($p < .001$), levator ($p = .008$), and zygomaticus ($p < .001$), but not frontalis ($p = .200$). The activity in the depressor was significantly lower than in the frontalis ($p < .001$), levator ($p = .001$), and zygomaticus ($p < .001$). The activity in the frontalis was significantly higher than in the levator ($p = .001$) and the zygomaticus ($p < .001$). The activity in the levator was not significantly different from the zygomaticus ($p = .817$) [frontalis = corrugator > levator = zygomaticus > depressor]. During the **BM** condition the activity was highest in the depressor, followed by the levator, zygomaticus, corrugator, and frontalis; the activity between the muscles differed significantly (p's < .005), although for the difference in activity between the corrugator and frontalis a trend was found ($p = .052$). This shows that the muscles surrounding the mouth were activated by the pen holding during the BM condition. Additionally, the EMG activity was compared between the four experimental conditions for each face muscle site. The pattern of activity found for the depressor, levator, zygomaticus, and corrugator was: IM > BM > FER = SM, but for the frontalis: IM > BM = SM = FER ('>' significantly greater, $p < .05$; '=' not significant). That is, the frontalis was uninfluenced by the pen holding during the BM condition, but increased activity was induced in all other muscles by this experimental manipulation.

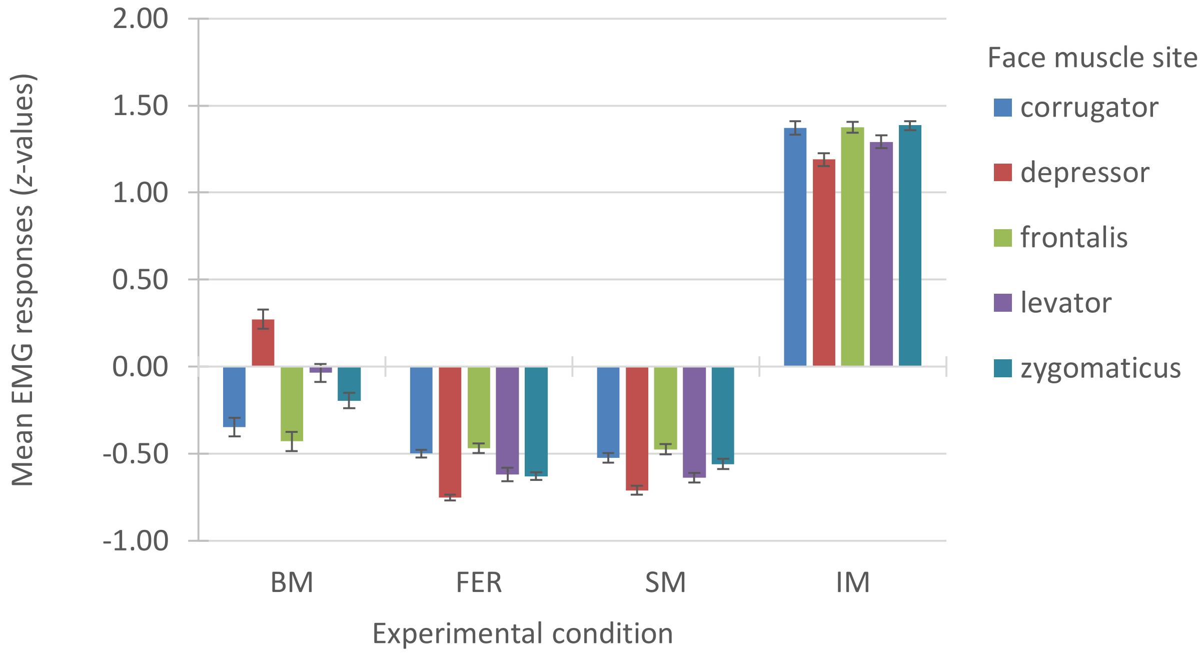


Figure 24. Mean EMG responses for each of the four experimental conditions across the five face muscle sites. Since the data were *z*-standardised (across experimental conditions, within-subject, within-muscle), a mean of 0 does not reflect an absence of activity, but the average activity across experimental conditions. BM = blocked mimicry, FER = no face movement manipulation with facial emotion recognition, SM = no face movement manipulation (without facial emotion recognition), IM = explicit imitation. Error bars represent standard errors of the means.

Aim 1: Emotion-specific EMG patterns.

Emotion-specific EMG patterns during the SM condition.

Emotion-specific EMG patterns during the SM condition were investigated with a repeated measures ANOVA with face muscle sites (5) and emotions (10) as within-subject factor. Due to the *z*-standardisation within face muscle sites, a main effect of face muscle site could not be computed, since all muscles had a mean of 0 and *SD* of 1. The main effect of *emotion* was not significant ($F(9,747) = 0.68$, $p = .725$, partial $\eta^2 = .008$, power = .34). Due to violation of the assumption of sphericity for the emotion*muscle interaction ($W(665) = .00$, $p < .001$), Greenhouse-Geisser adjustment of degrees of freedom was applied. The interaction of *muscle*emotion* was significant ($F(21.97,1823.51) = 6.12$, $p < .001$, partial $\eta^2 = .069$, power = 1.00). Pairwise comparisons showed that in response to *anger* faces the EMG activity was significantly higher in the corrugator than in the zygomaticus ($p < .001$), depressor ($p = .002$), and frontalis ($p = .002$), and levator ($p = .007$). A trend was also found for the difference in activity between the depressor and frontalis ($p = .074$). There were no other significant differences in EMG activity between muscles (p's > .330). [corrugator > depressor = levator = frontalis = zygomaticus]; see Figure 25.

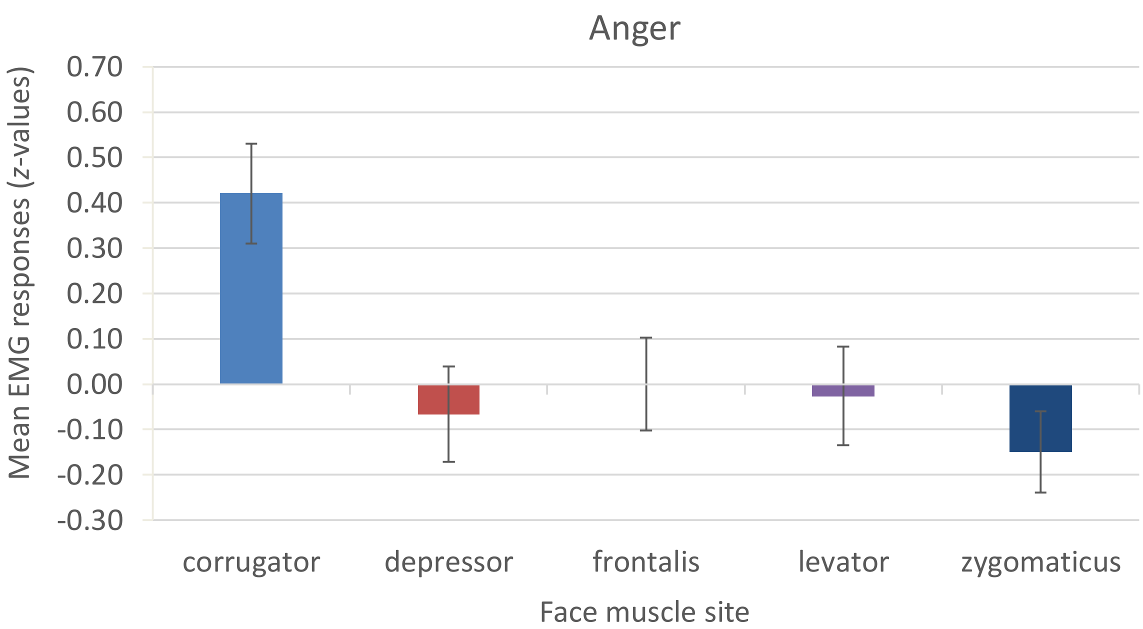


Figure 25. Facial mimicry responses for anger across the five face muscle sites. An increase in corrugator activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

There were no significant differences in EMG activity between face muscles in response to *disgust* faces (*p*'s > .754), although a trend was found for the activity between corrugator and zygomaticus (*p* = .075). [(corrugator = depressor = levator = frontalis) = zygomaticus], see Figure 26.

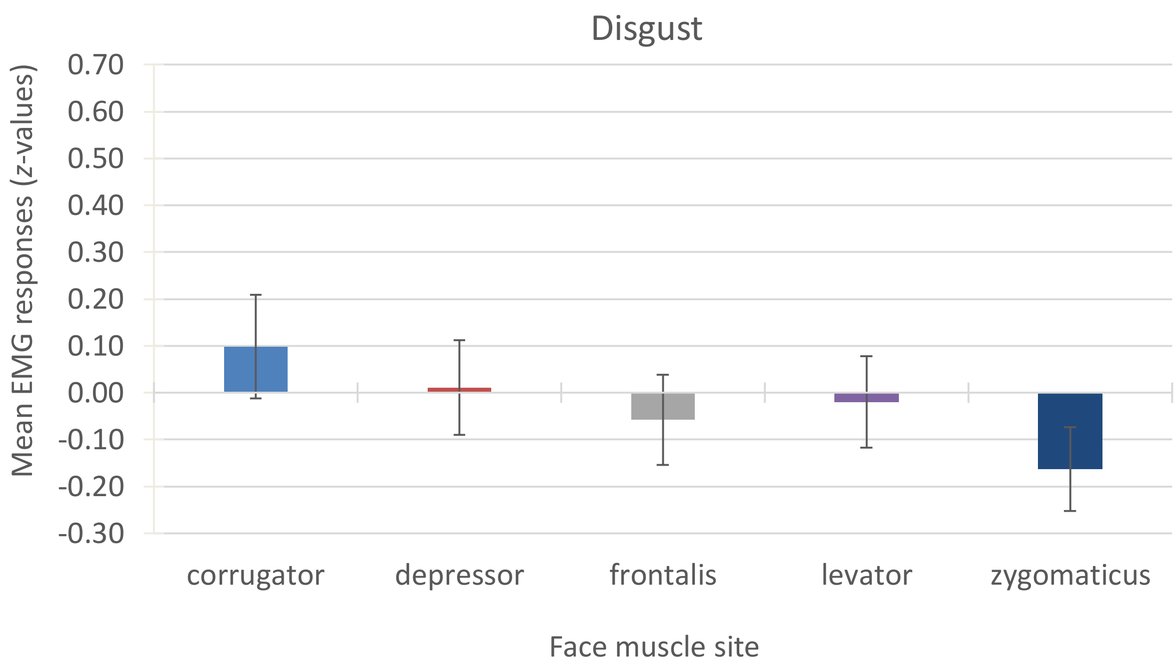


Figure 26. Facial mimicry responses for disgust across the five face muscle sites. An increase in levator activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *fear* faces, the activity in the zygomaticus was significantly higher than in the depressor (p = .049). There were no other significant differences between muscles in response to fear (p's > .086). [(zygomaticus = levator = corrugator = frontalis =) > depressor], see Figure 27.

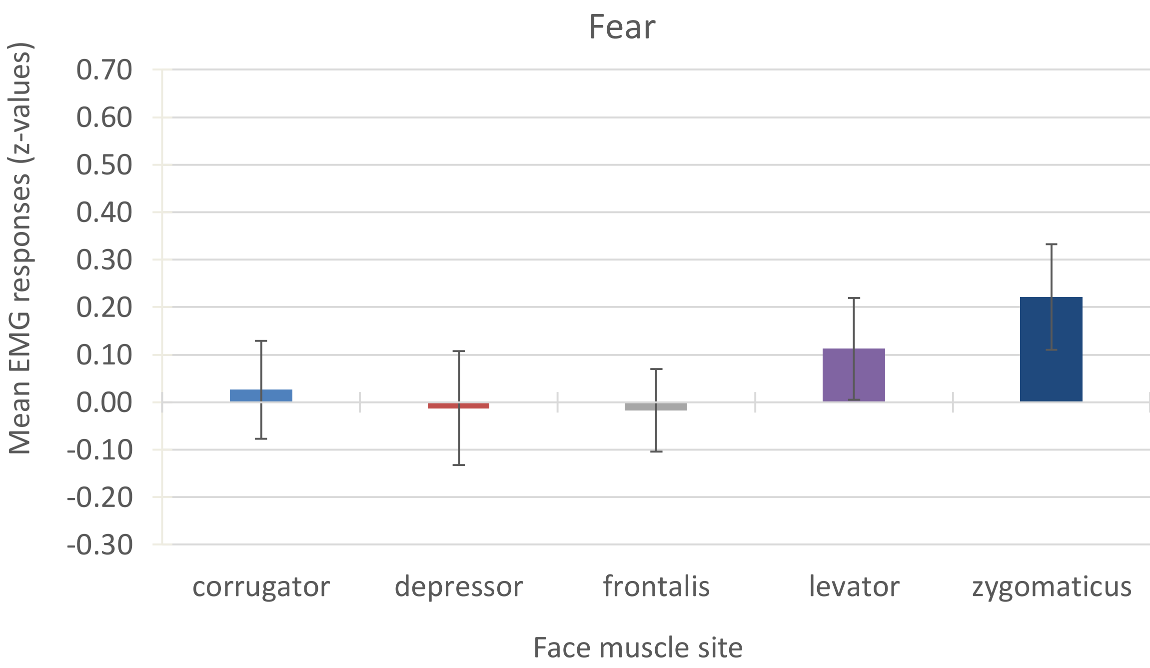


Figure 27. Facial mimicry responses for fear across the five face muscle sites. An increase in corrugator activity and lateral frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *sadness* faces, the EMG activity was significantly higher in corrugator than depressor ($p < .001$), levator ($p = .002$), frontalis ($p = .004$), and zygomaticus ($p = .001$). There were no other significant differences in EMG activity between muscles (p's > .198). [corrugator > frontalis = levator = zygomaticus = depressor], see Figure 28.

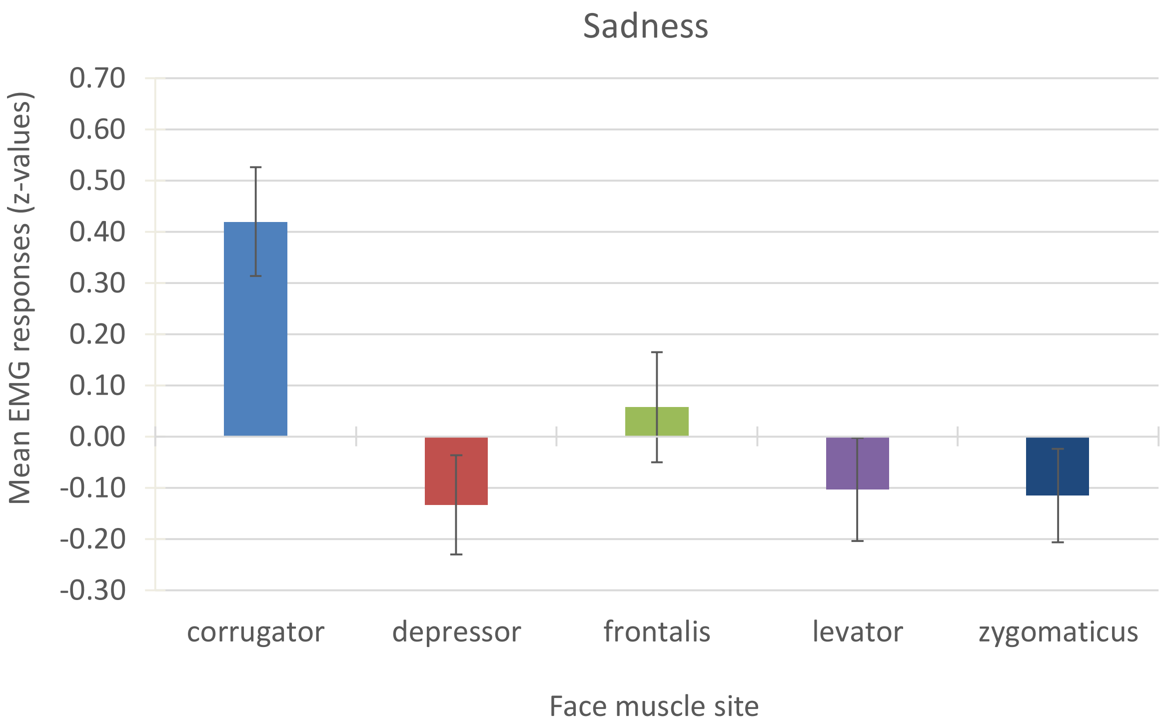


Figure 28. Facial mimicry responses for sadness across the five face muscle sites. An increase in corrugator activity and depressor activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *surprise* faces, the EMG activity was significantly higher in frontalis than corrugator ($p < .001$), depressor ($p = .001$), levator ($p < .001$), and zygomaticus ($p = .003$). There were no other significant differences in EMG activity between muscles (p's > .226). [frontalis > zygomaticus = depressor = levator = corrugator], see Figure 29.

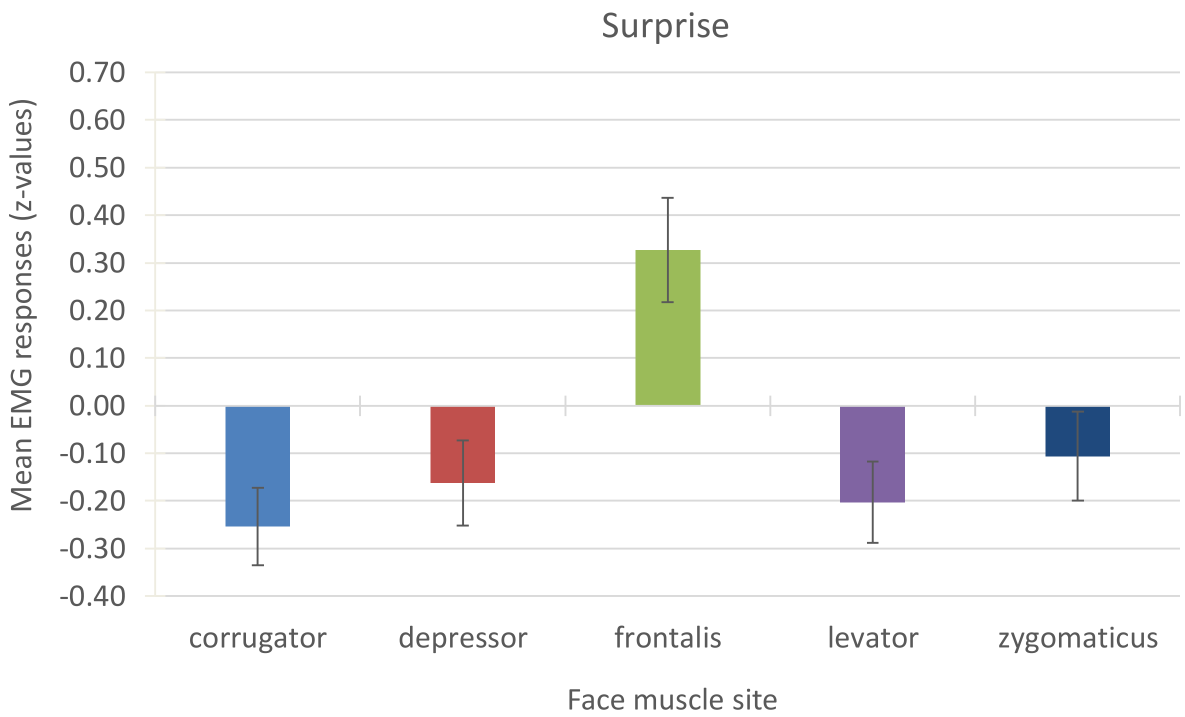


Figure 29. Facial mimicry responses for surprise across the five face muscle sites. An increase in frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *happiness* faces, the EMG activity was significantly higher in the zygomaticus than corrugator ($p = .002$), depressor ($p = .005$), and frontalis ($p < .001$). The activity in the levator was also significantly higher than in the frontalis ($p = .001$), corrugator ($p = .013$), and a trend was found in comparison to the depressor ($p = .056$). There were no other significant differences in EMG activity between muscles (p's = .132). [zygomaticus = levator > depressor = corrugator = frontalis], see Figure 30.

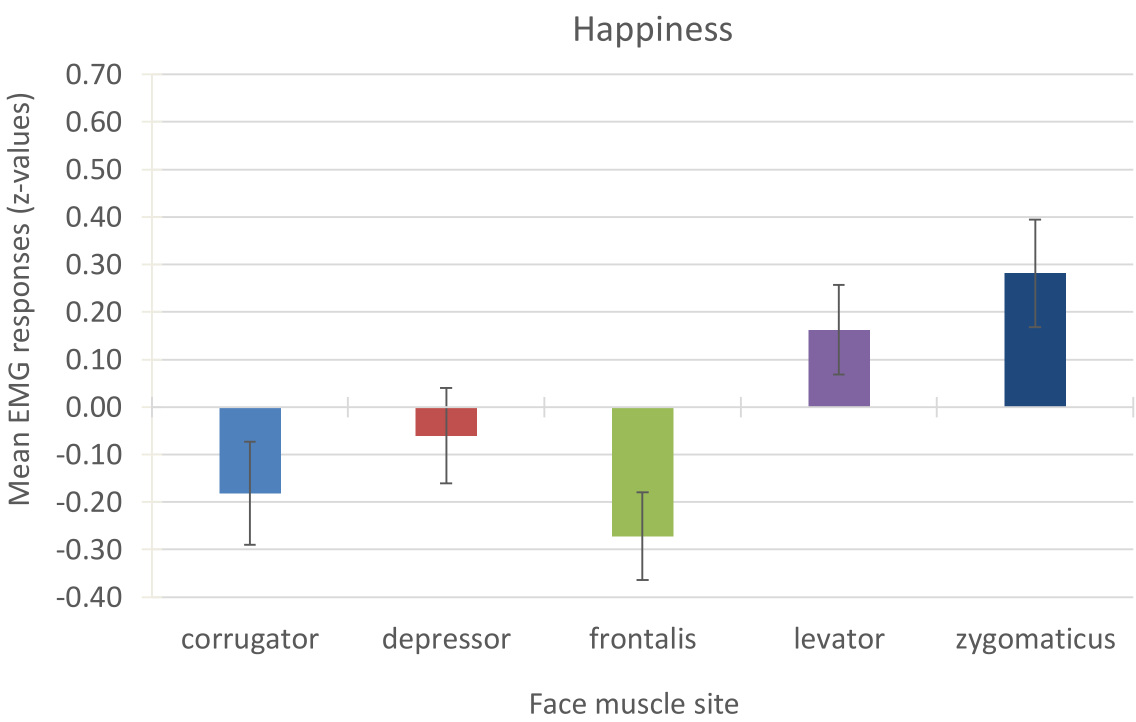


Figure 30. Facial mimicry responses for happiness across the five face muscle sites. An increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *neutral* faces, the EMG activity in the zygomaticus was significantly lower than in the corrugator (p = .003), depressor (p = .003), frontalis (p < .001), and levator (p = .002). The activity in the frontalis was significantly higher than in the depressor (p = .006) and a trend was found in comparison to the corrugator (p = .052). [(frontalis = levator =) > corrugator = depressor > zygomaticus], see Figure 31.

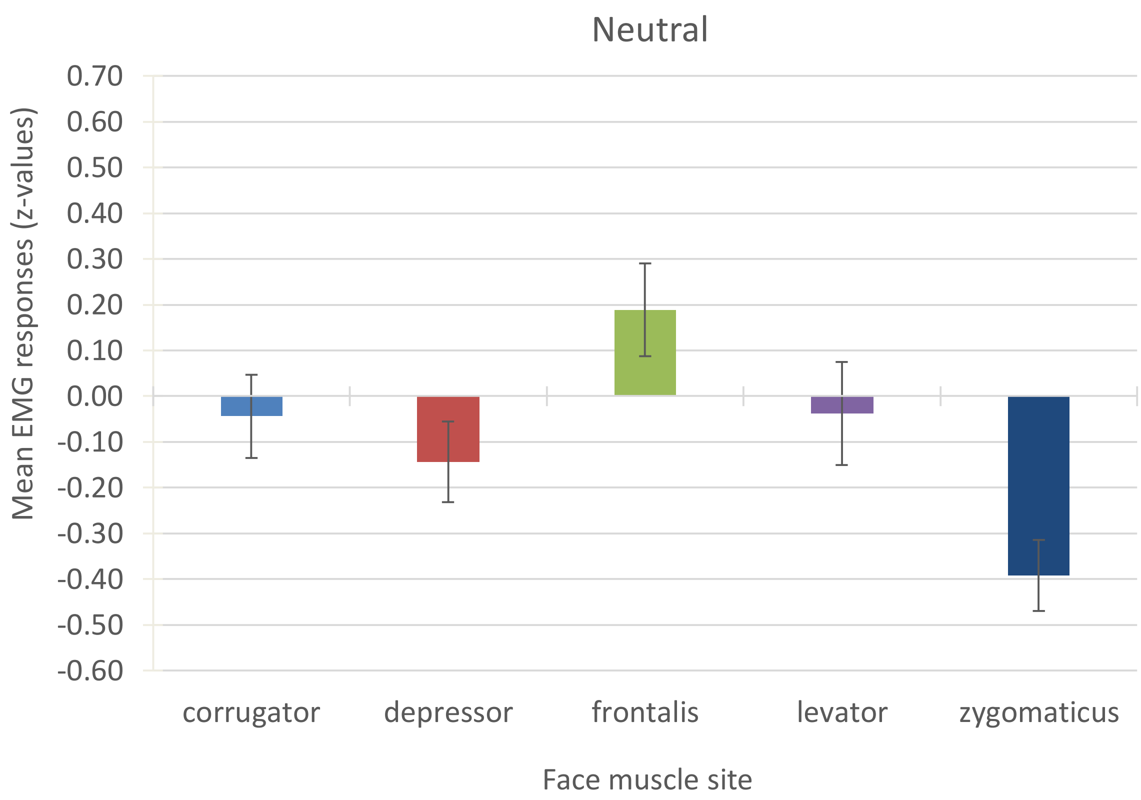


Figure 31. Facial mimicry responses for neutral across the five face muscle sites. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *contempt* faces, the EMG activity in the depressor was significantly higher than in the zygomaticus (p = .004) and levator (p = .045). There were no other significant differences in EMG activity between muscles (p's > .098). [(depressor = frontalis = corrugator = levator =) > zygomaticus], see Figure 32.

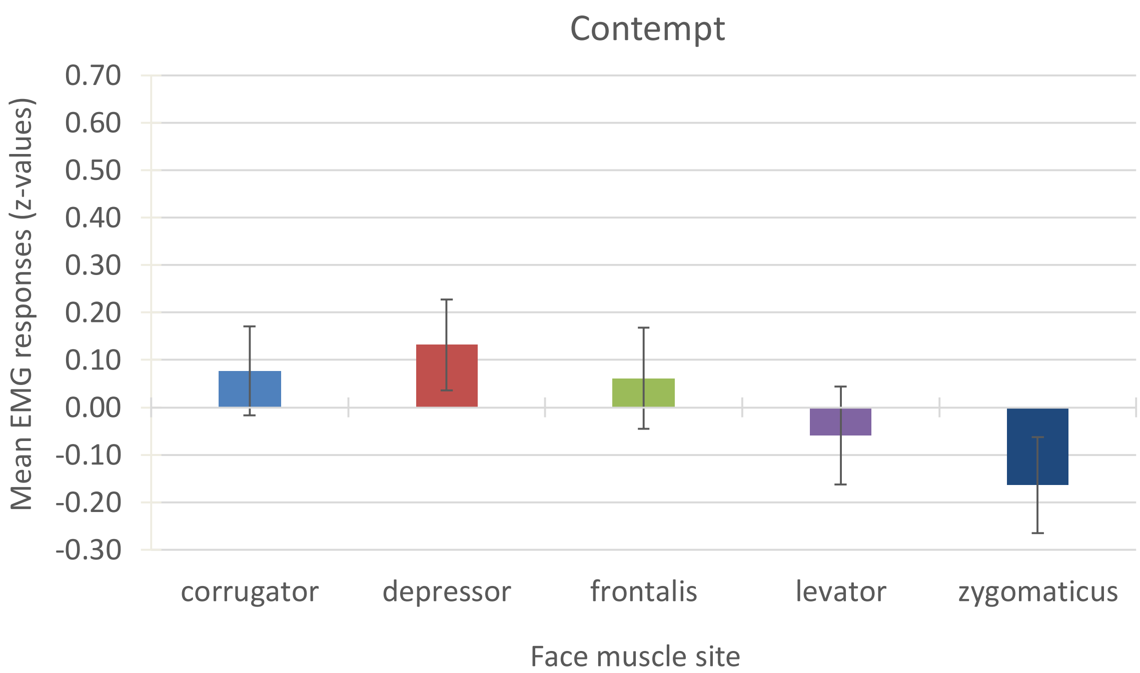


Figure 32. Facial mimicry responses for contempt across the five face muscle sites. An increase in zygomaticus and frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were z-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *embarrassment* faces, the EMG activity was significantly higher in the depressor than the zygomaticus (p = .033) and the levator (p = .031). There were no other significant differences in EMG activity between muscles in response to embarrassment (p's > .125). [(depressor = frontalis = corrugator =) > zygomaticus = levator], see Figure 33.

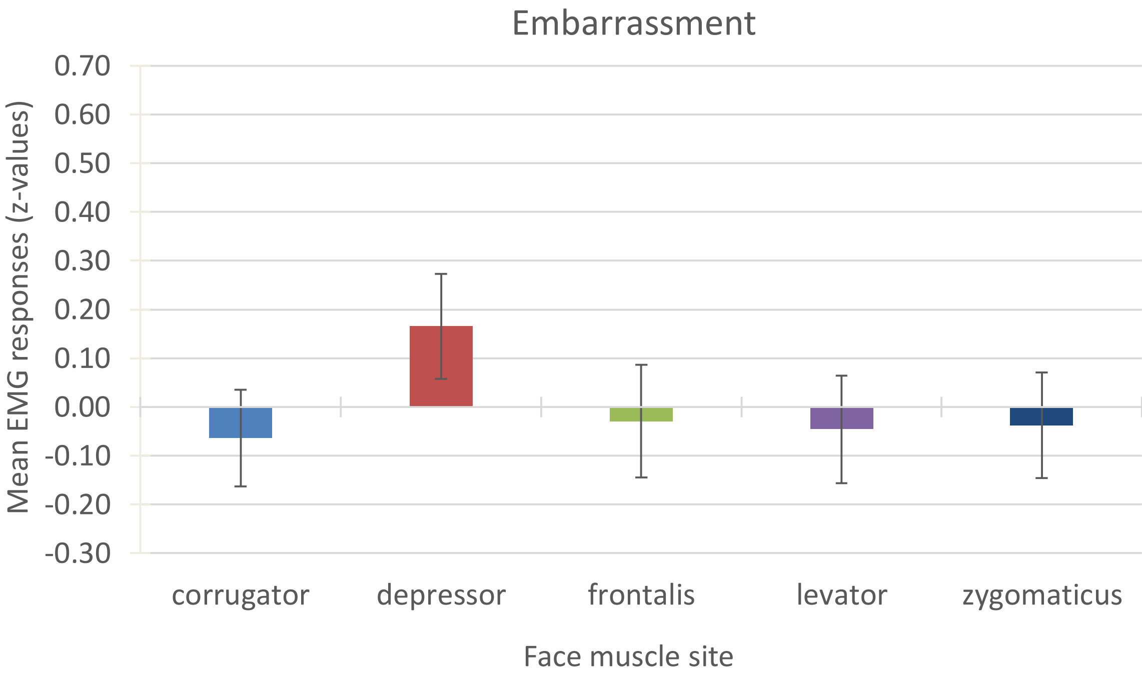


Figure 33. Facial mimicry responses for embarrassment across the five face muscle sites. An increase in depressor activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were z-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

In response to *pride* faces, the EMG activity in zygomaticus was significantly higher than in corrugator ($p < .001$), depressor ($p = .002$), frontalis ($p < .001$), and levator ($p = .001$). The EMG activity was also significantly higher in levator than corrugator ($p < .001$) and frontalis ($p = .002$), and in depressor higher than in corrugator and frontalis (p's $< .001$). In the frontalis, the activity was significantly higher than in the corrugator ($p = .049$). The difference between levator and depressor was not significant ($p = .638$). [zygomaticus > depressor = levator > frontalis > corrugator], see Figure 34.

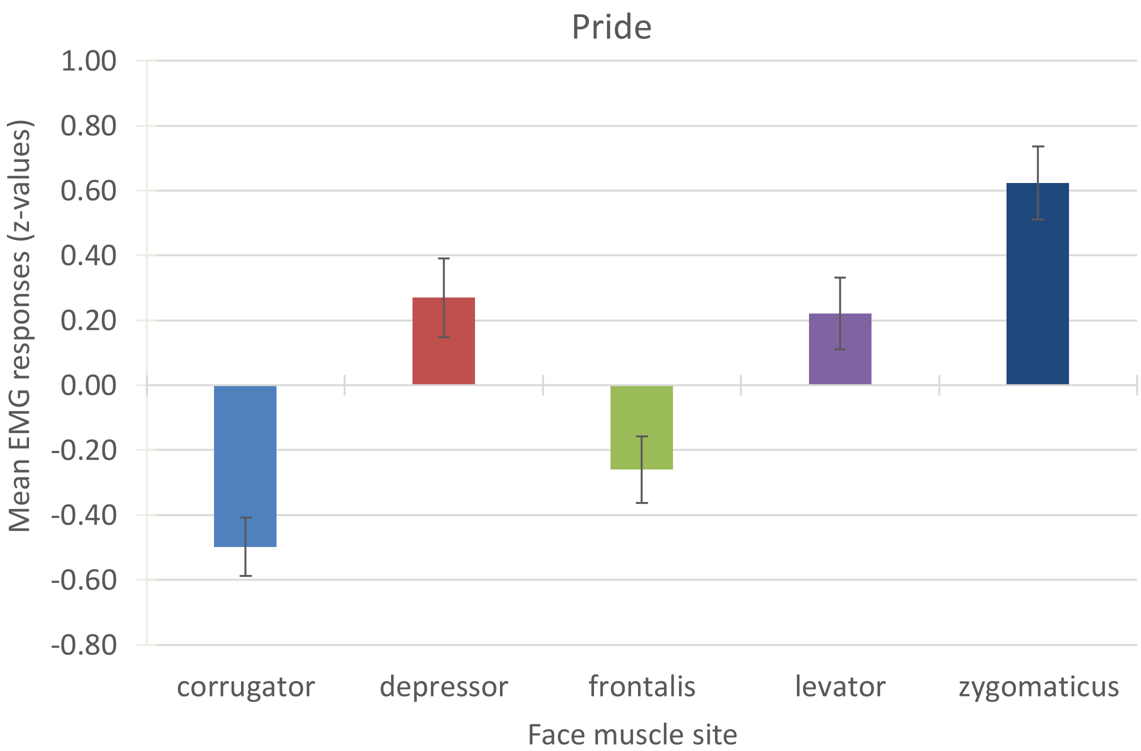


Figure 34. Facial mimicry responses for pride across the five face muscle sites. An increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

Emotion-specific EMG patterns during the IM condition.

Emotion-specific EMG patterns during the IM condition were investigated with a repeated measures ANOVA with face muscle sites (5) and emotions (10) as within-subject factors. Due to violation of the assumption of sphericity for emotion ($W(44) = .15$, $p < .001$) and the emotion*muscle interaction ($W(665) = .00$, $p < .001$), Greenhouse-Geisser adjustments of degrees of freedom were applied. Due to the standardisation within muscle sites, a main effect of muscle could not be computed. The main effect of *emotion* was significant ($F(6.83,566,85) = 131.17$, $p < .001$, partial $\eta^2 = .612$, power = 1.00). Pairwise comparisons showed that the EMG responses of the majority of the emotions were significantly different from each other (*p*'s < .05), except anger was not significantly different from happiness ($p = .380$) and sadness ($p = .263$), contempt was not significantly different from surprise ($p = .365$), embarrassment was not significantly different from sadness ($p = .369$), and fear was not significantly different from pride ($p = .987$). Trends were found for the differences in EMG responses between anger and embarrassment ($p = .067$) and sadness and happiness ($p = .081$). The following order of EMG responses across all muscle sites was found: disgust > fear = pride > (happiness = anger =) > sadness = embarrassment > surprise = contempt > neutral; see Figure 35.

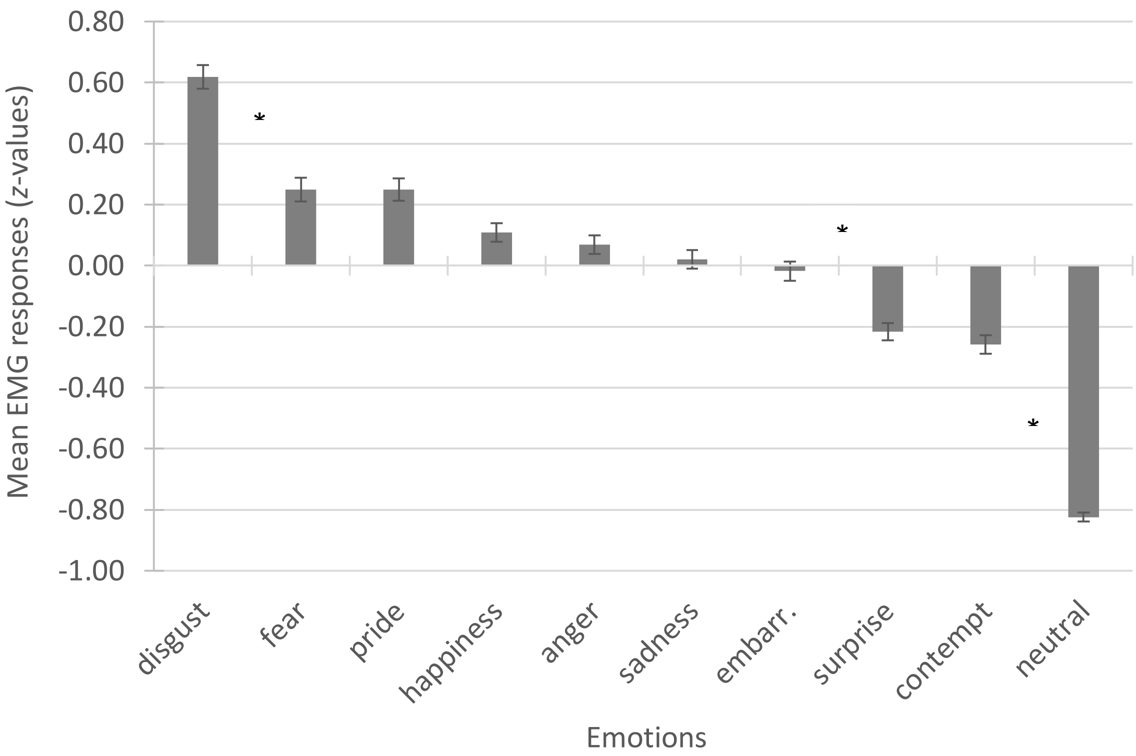


Figure 35. EMG responses during explicit imitation per emotion over the five face muscle sites combined. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the muscles combined across emotions. Embarr. = embarrassment. Error bars represent standard errors of the means. *significant difference between emotions, $p < .05$.

The interaction of *muscle*emotion* was significant (F(12.89,1069.99) = 157.41, p < .001, partial η^2 = .655, power = 1.00). Pairwise comparisons showed that during explicit imitation of *anger* faces, the EMG activity was significantly higher in the corrugator than in the zygomaticus (p < .001), depressor (p < .001), levator (p < .001), and frontalis (p < .001). The EMG activity was also significantly higher in the frontalis than the depressor (p < .001), levator (p < .001), and zygomaticus (p < .001), and in the levator higher than in the zygomaticus (p = .003). There were no other significant differences in EMG activity between muscles (p's > .287). [corrugator > frontalis > levator = depressor = zygomaticus], see Figure 36.

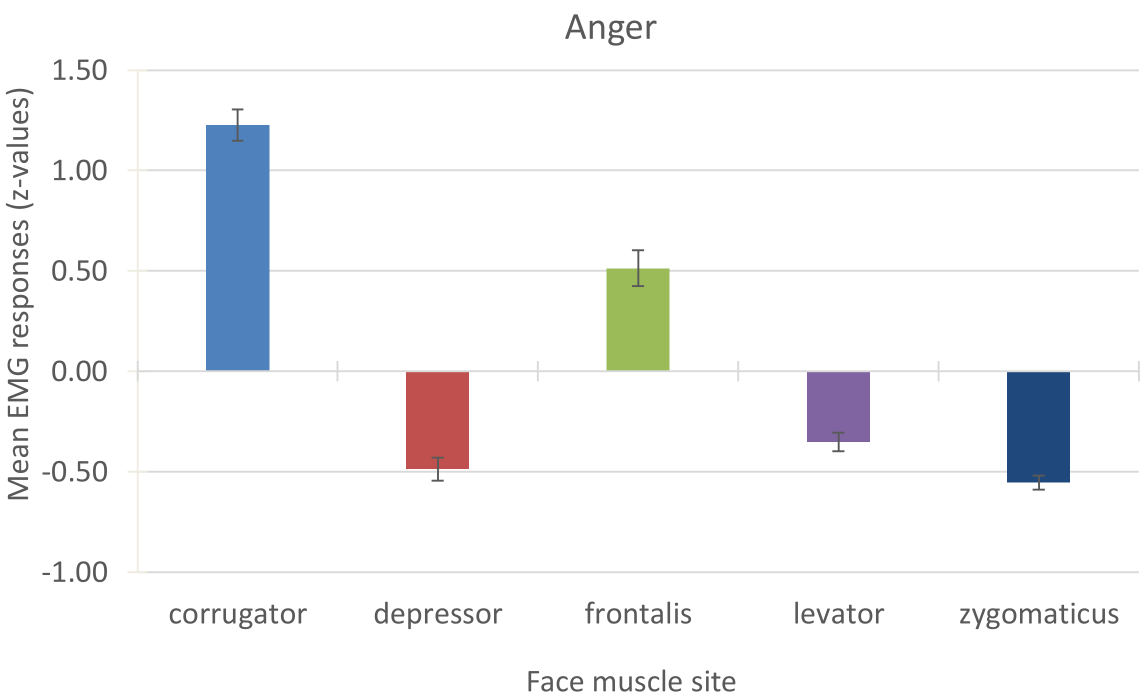


Figure 36. EMG responses during explicit anger imitation across the five face muscle sites. An increase in corrugator activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

During the imitation of *disgust* faces, the EMG activity was significantly higher in the levator than the depressor ($p < .001$), the frontalis ($p < .001$), and the zygomaticus ($p < .001$). The EMG activity was also significantly higher in the corrugator than the depressor ($p < .001$), frontalis ($p < .001$), and zygomaticus ($p < .001$), and higher in the frontalis than the depressor ($p = .012$) and the zygomaticus ($p < .001$). There were no other significant differences in EMG activity between muscles (p's $> .116$). [levator = corrugator > frontalis > depressor = zygomaticus], see Figure 37.

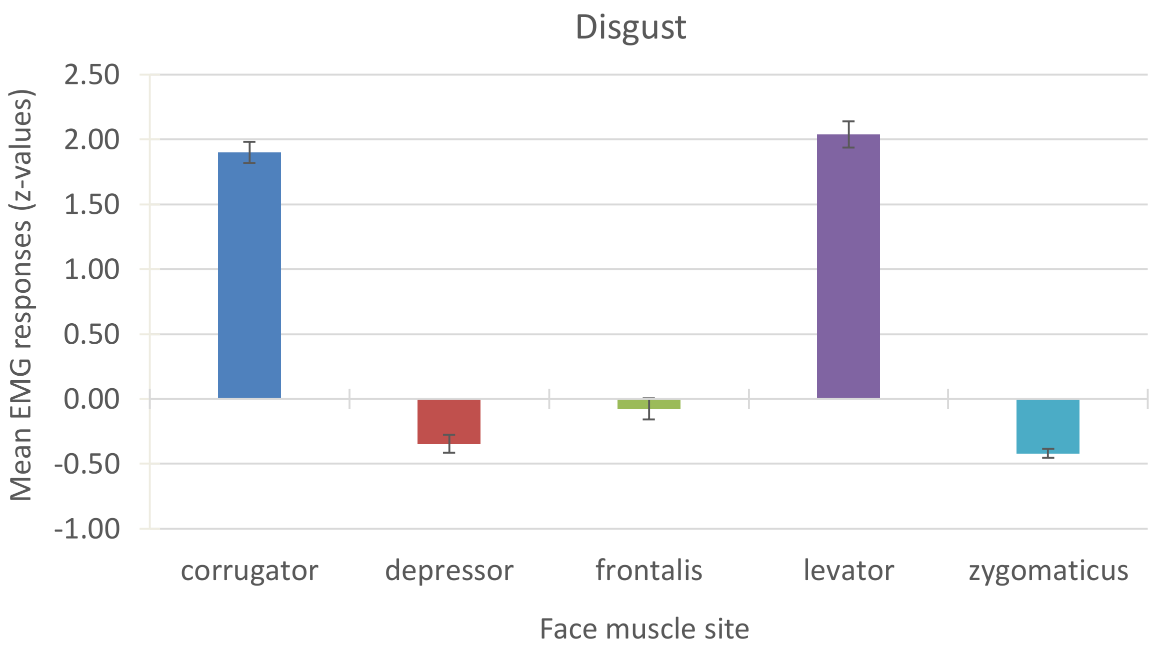


Figure 37. EMG responses during explicit disgust imitation across the five face muscle sites. An increase in levator activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

When imitating *fear* faces, the EMG activity was significantly higher in the depressor than the corrugator ($p < .001$), levator ($p < .001$), and zygomaticus ($p < .001$). The EMG activity was also significantly higher in the frontalis than the corrugator ($p < .001$), the levator ($p < .001$), and zygomaticus ($p < .001$), and higher in the corrugator than the zygomaticus ($p = .002$). There were no other significant differences in EMG activity between muscles (p's > .109). [frontalis = depressor > corrugator > levator = zygomaticus], see Figure 38.

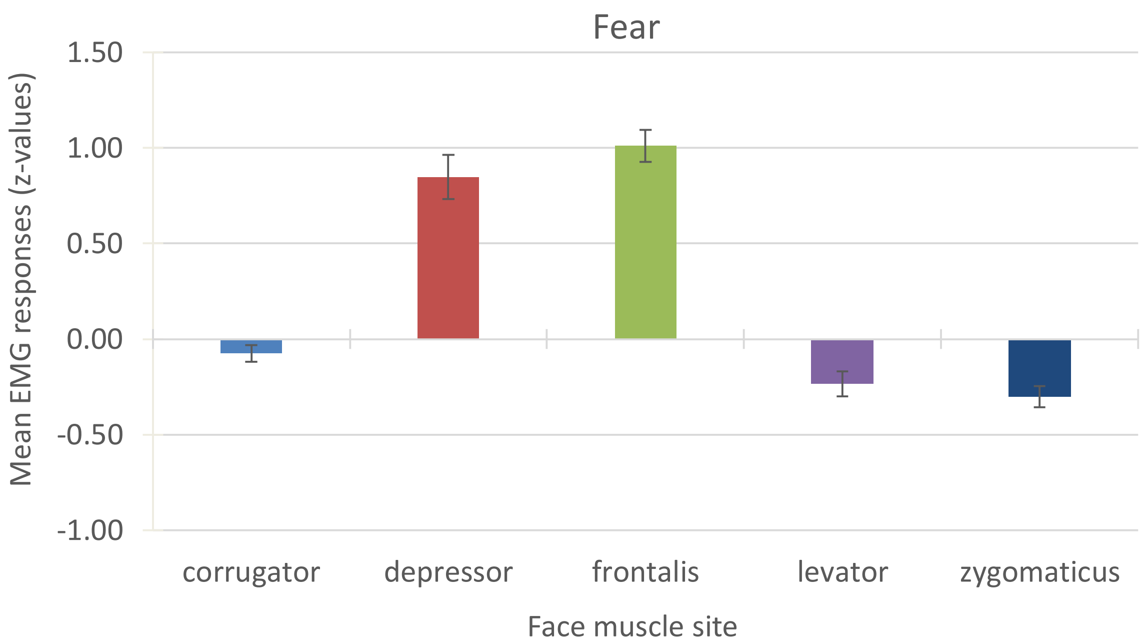


Figure 38. EMG responses during explicit fear imitation across the five face muscle sites. An increase in corrugator activity and frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

When imitating *sadness* faces, the EMG activity was significantly higher in the corrugator than the levator ($p < .001$) and zygomaticus ($p < .001$). The EMG activity was also significantly higher in the depressor than the levator ($p < .001$) and zygomaticus ($p < .001$), and higher in the frontalis than the levator ($p < .001$) and zygomaticus ($p < .001$). There were no other significant differences in EMG activity between muscles (p's = 1.00). [corrugator = depressor = frontalis > levator > zygomaticus], see Figure 39.

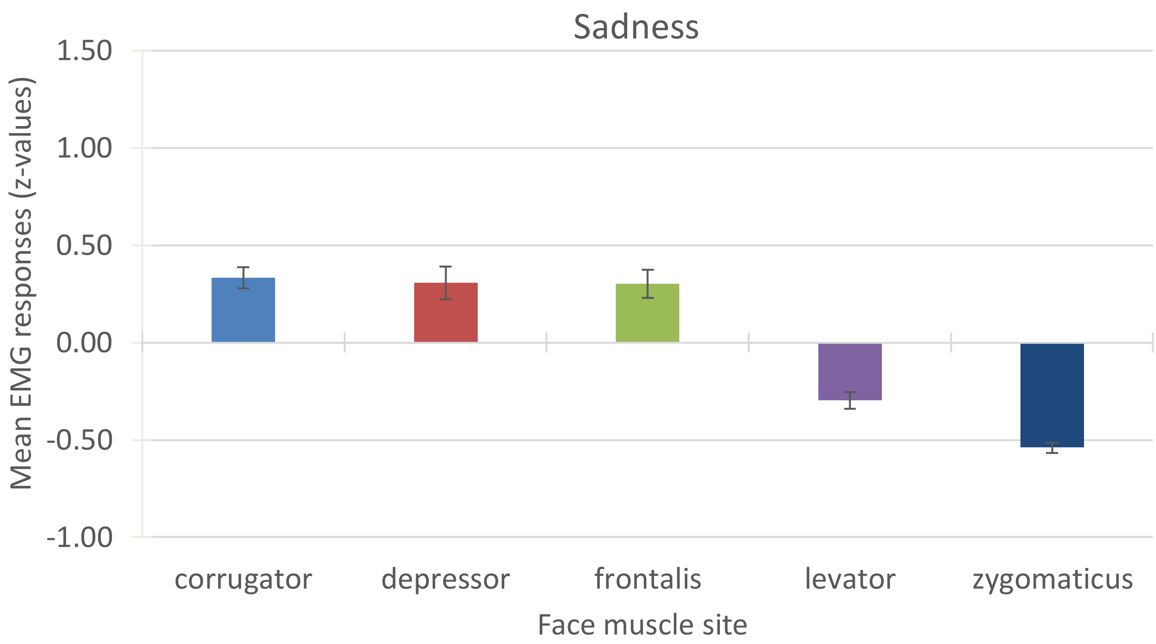


Figure 39. EMG responses during explicit sadness imitation across the five face muscle sites. An increase in corrugator activity and depressor activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

During the imitation of *surprise* faces, the EMG activity was significantly higher in the frontalis than the corrugator ($p < .001$), the depressor ($p < .001$), the levator ($p < .001$), and the zygomaticus ($p < .001$). The EMG activity was also significantly higher in the corrugator than the levator ($p < .001$) and zygomaticus ($p < .001$), and higher in the depressor than the levator ($p < .001$) and zygomaticus ($p < .026$). There were no other significant differences in EMG activity between muscles (p's = 1.00). [frontalis > depressor = corrugator > zygomaticus = levator], see Figure 40.

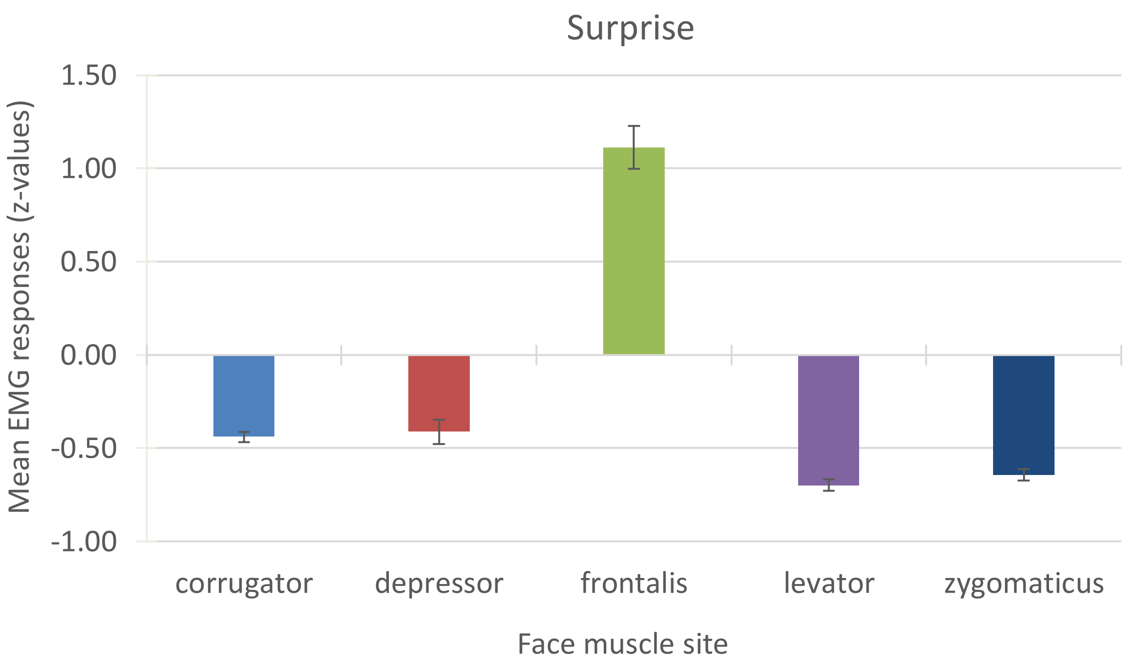


Figure 40. EMG responses during explicit surprise imitation across the five face muscle sites. An increase in frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

During the imitation of *happiness* faces, the EMG activity was significantly higher in the zygomaticus than the corrugator ($p < .001$), the depressor ($p < .001$), the levator ($p < .001$), and the frontalis ($p < .001$). The EMG activity was also significantly higher in the levator than the corrugator ($p < .001$) and frontalis ($p < .001$), higher in the depressor than the corrugator ($p < .001$) and frontalis ($p < .001$), and higher in the corrugator than the frontalis ($p = .002$). There were no other significant differences in EMG activity between muscles (*p*'s = 1.00). [zygomaticus > depressor = levator > corrugator > frontalis], see Figure 41.

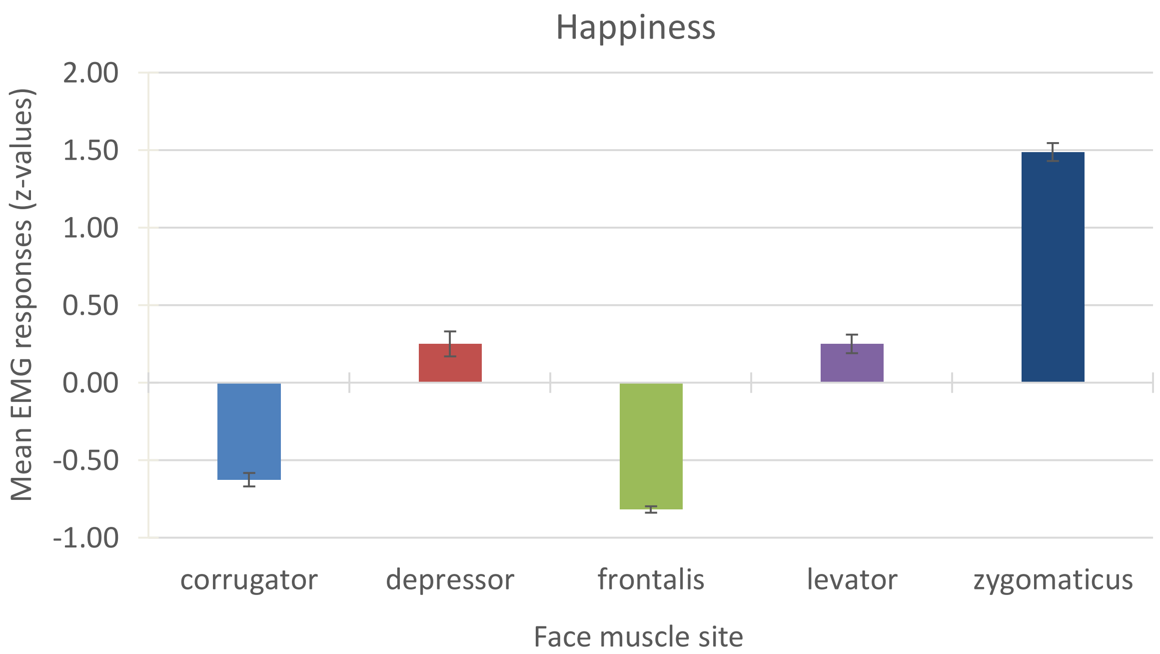


Figure 41. EMG responses during explicit happiness imitation across the five face muscle sites. An increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

When presented with *neutral* faces, the EMG activity was significantly higher in the corrugator than the depressor ($p < .001$), the levator ($p < .001$), the frontalis ($p = .020$), and the zygomaticus ($p = .008$). The EMG activity was also significantly higher in the frontalis than the depressor ($p < .001$), higher in the zygomaticus than the depressor ($p < .001$) and the levator ($p = .001$), and higher in the levator than the depressor ($p < .001$). A trend was found for the EMG difference between the frontalis and the levator ($p = .050$). The difference between zygomaticus and frontalis was not significant ($p = .645$). [corrugator > frontalis = zygomaticus > levator > depressor], see Figure 42.

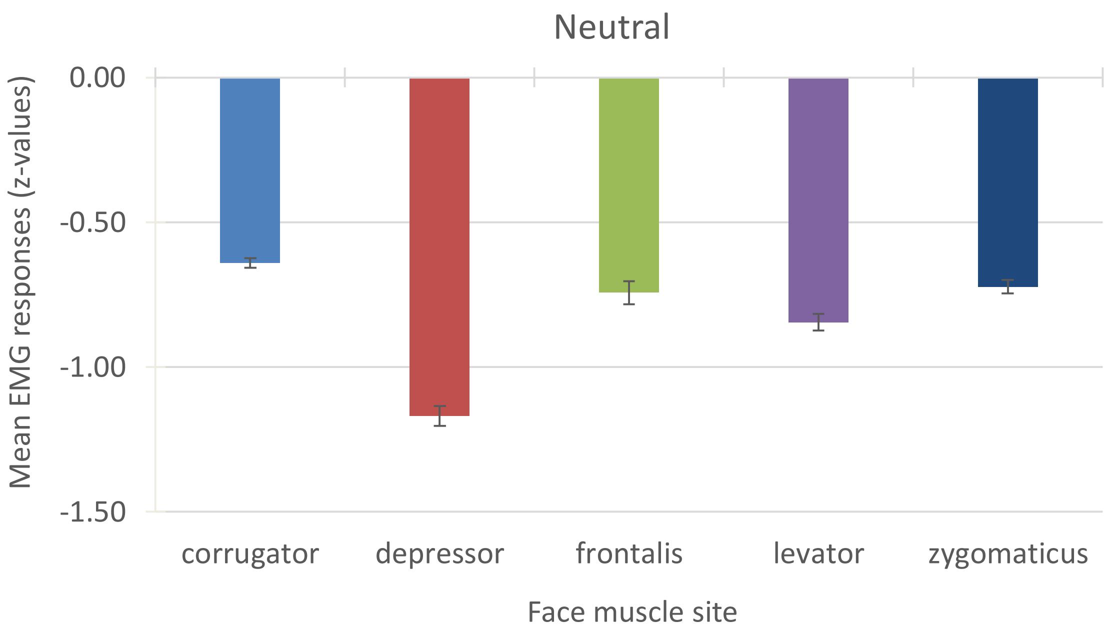


Figure 42. EMG responses during explicit neutral faces imitation across the five face muscle sites. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

When imitating *contempt* faces, the EMG activity was significantly higher in the frontalis than the corrugator ($p < .001$), the levator ($p < .001$), the depressor (p = .019), and the zygomaticus (p = .045). The EMG activity was also significantly higher in the zygomaticus than the corrugator ($p < .001$) and the levator ($p = .001$). The EMG activity was significantly higher in the depressor than the corrugator ($p = .012$). There was a trend for higher EMG activity in the levator than the corrugator ($p = .082$). There were no other significant differences in EMG activity between muscles (p's > .093). [frontalis > zygomaticus = depressor = levator > corrugator], see Figure 43.

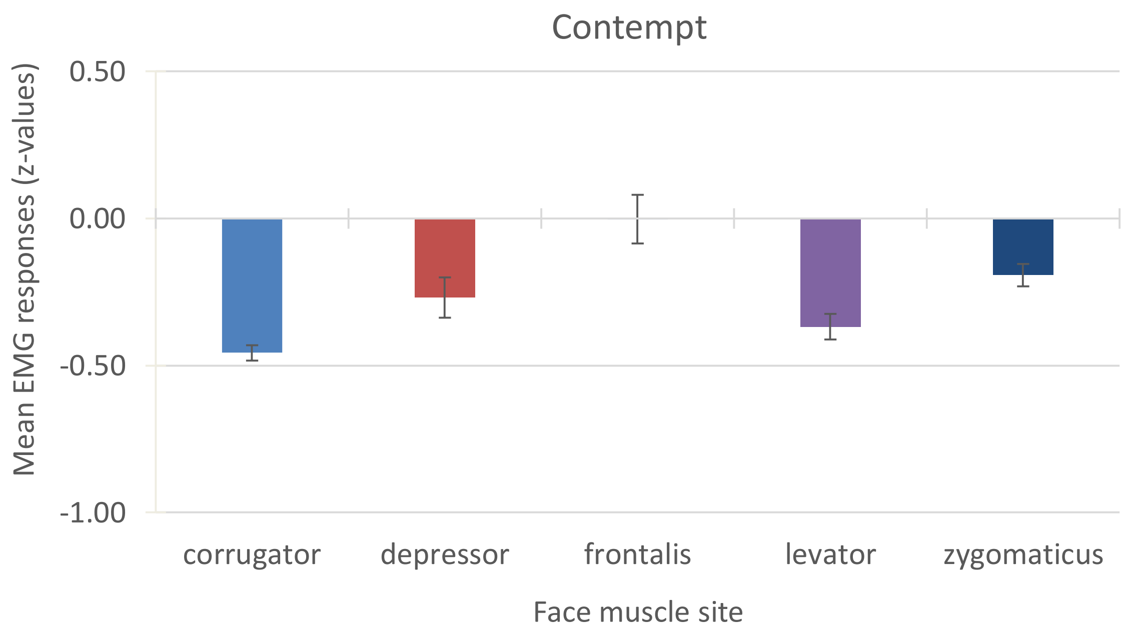


Figure 43. EMG responses across the five face muscle sites during explicit imitation of contempt. An increase in zygomaticus and frontalis activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

During the imitation of *embarrassment* faces, the EMG activity was significantly higher in the depressor than the corrugator ($p < .001$), the frontalis ($p < .001$), the levator ($p < .001$), and the zygomaticus ($p < .001$). The EMG activity was also significantly higher in the levator than the corrugator ($p < .001$) and frontalis ($p < .001$), and higher in the zygomaticus than the corrugator ($p < .001$) and frontalis ($p < .001$). There were no other significant differences in EMG activity between muscles (p's > .469). [depressor > zygomaticus = levator > corrugator = frontalis], see Figure 44.

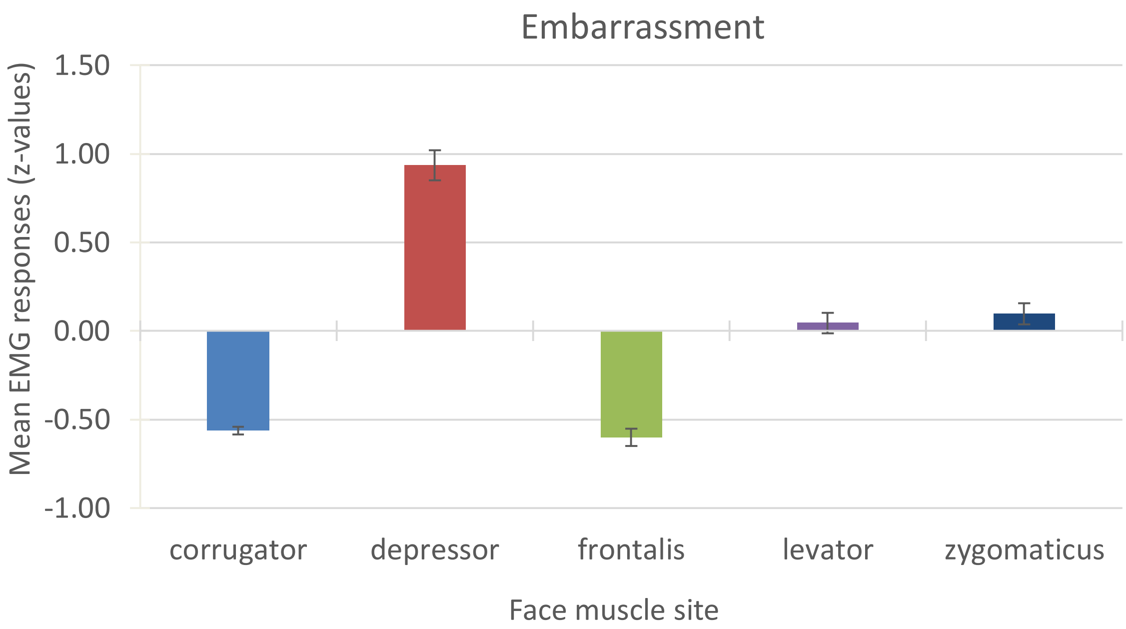


Figure 44. EMG responses across the five face muscle sites during explicit imitation of embarrassment. An increase in depressor activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

During the imitation of *pride* faces, the EMG activity was significantly higher in the zygomaticus than the corrugator ($p < .001$), the depressor ($p < .001$), the levator ($p < .001$), and the frontalis ($p < .001$). The EMG activity was also significantly higher in the depressor than the corrugator ($p < .001$) and the frontalis ($p < .001$), and higher in the levator than the frontalis ($p < .001$) and the corrugator (p < .001). There were no other significant differences in EMG activity between muscles (p's > .222). [zygomaticus > levator = depressor > corrugator = frontalis], see Figure 45.

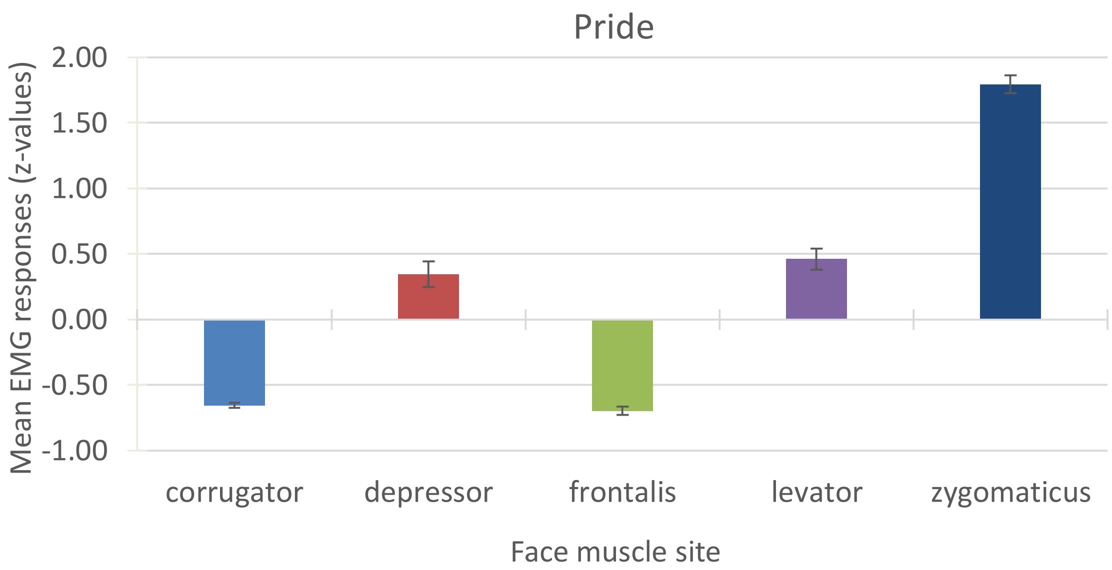


Figure 45. EMG responses across the five face muscle sites during explicit imitation of pride. An increase in zygomaticus activity compared to baseline was expected that would be higher than the increase in activity in the other muscles. Since the data were *z*-standardised (within: experimental conditions, subject, muscle), a mean of 0 does not reflect an absence of activity, but the average activity of the respective muscle across emotions. Error bars represent standard errors of the means.

Aim 2: Intensity of observed facial expression and intensity of EMG responses.

Intensity of observed facial expression and intensity of EMG responses during explicit imitation.

The results of the 5 face muscle sites (corrugator, depressor, frontalis, levator, zygomaticus) x 4 intensities (high, intermediate, low, neutral) repeated measures ANOVA from the IM condition showed that Sphericity was violated for the intensities ($W(5) = .19$, $p < .001$), face muscle sites ($W(9) = .65$, $p < .001$), and the interaction muscles*intensities ($W(77) = .00$, $p < .001$), hence, Greenhouse-Geisser adjustment of degrees of freedom was applied. The main effect of *intensity* was significant ($F(1.52,126.12) = 317.64$, $p < .001$, partial $\eta^2 = .793$, power = 1.00). Pairwise comparisons showed that the EMG responses in all four levels were significantly different from each other, in the order from high to low: high, intermediate, low, neutral (p's < .001), see Figure 46.

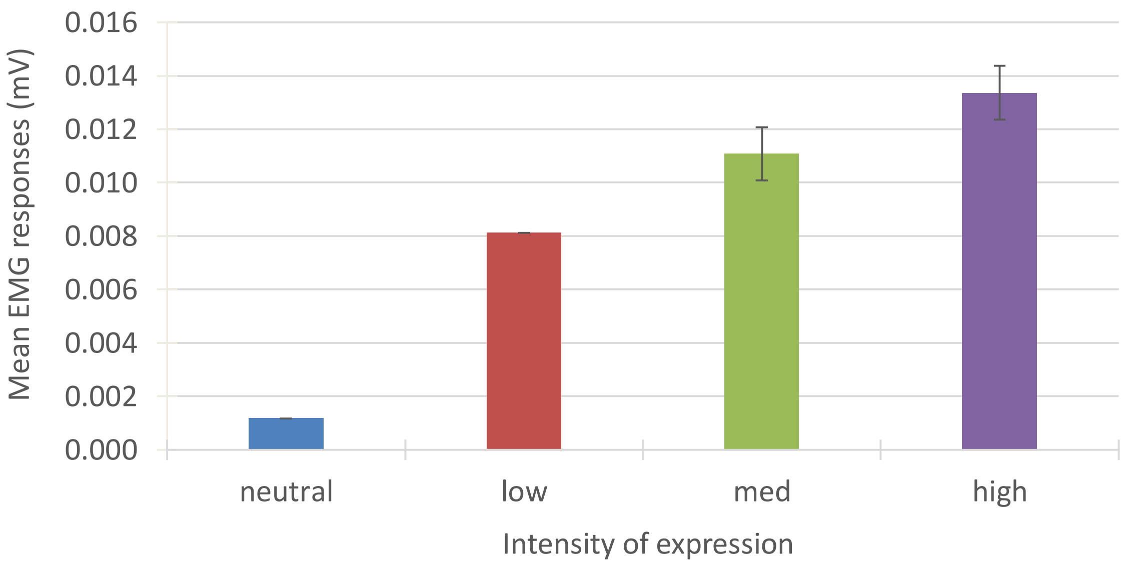


Figure 46. Face EMG responses at each intensity level during explicit imitation. Error bars represent the standard errors of the means.

The main effect of *face muscle site* was also significant ($F(3.26,270.53) = 40.56$, $p < .001$, partial $\eta^2 = .328$, power = 1.00). Pairwise comparisons showed that the highest EMG responses occurred in the depressor, followed by the corrugator, frontalis, zygomaticus, and levator. A trend was found for the differences between the zygomaticus and the levator ($p = .081$) and the zygomaticus and the frontalis ($p = .080$). All other comparisons between the face muscle sites in EMG responses were significant ($p < .001$), see Figure 47.

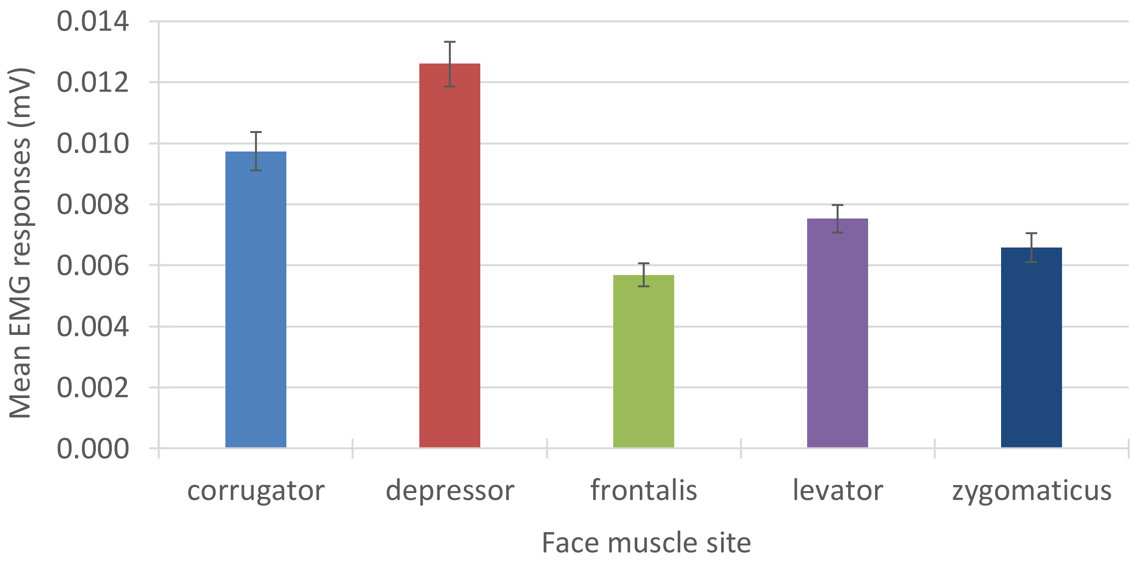


Figure 47. Face EMG responses for each muscle during explicit imitation. Error bars represent the standard errors of the means.

The interaction *intensity*face muscle site* was significant ($F(6.06,503.27) = 23.58$, $p < .001$, partial $\eta^2 = .221$, power = 1.00). Pairwise comparisons showed that within each face muscle site the EMG activity followed the order of lowest activity during explicit imitation of neutral faces through to highest activity during explicit imitation of high intensity facial expressions of emotion (p's ≤ .001), but with varying increases in EMG responses according to intensity level between the face muscle sites (see Figure 48).

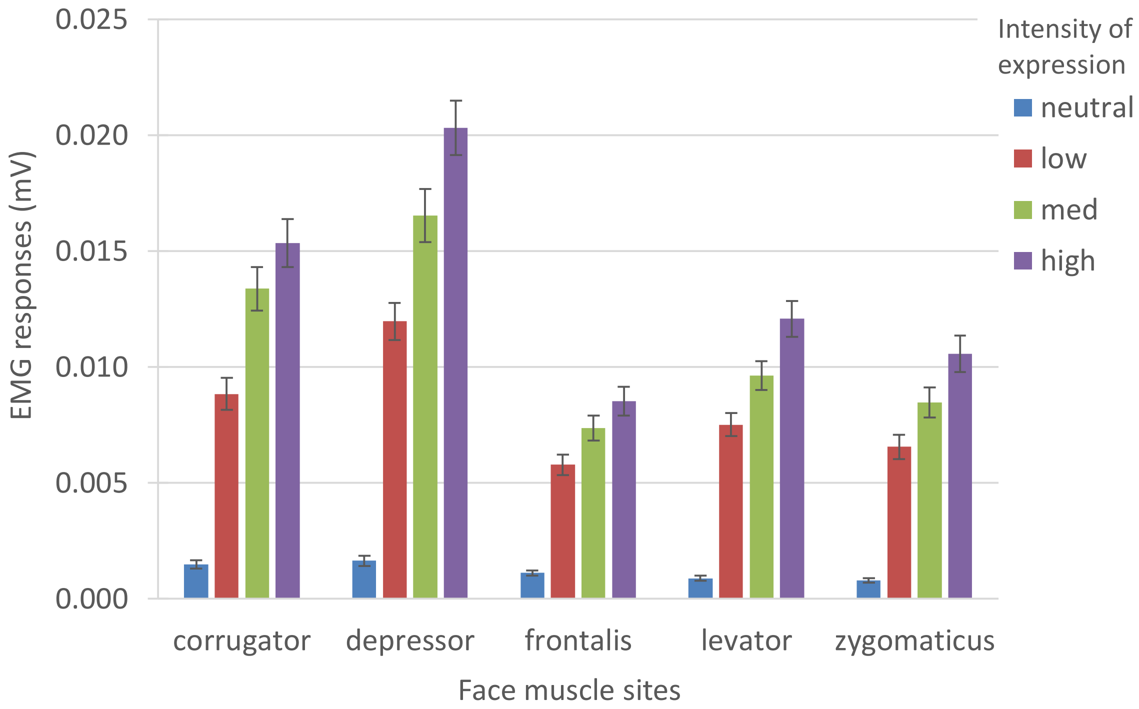


Figure 48. Face EMG responses across intensity levels for each muscle during explicit imitation. Error bars represent the standard errors of the means.

Intensity of observed facial expression and intensity of EMG responses during facial mimicry.

The results of the 5 face muscle sites (corrugator, depressor, frontalis, levator, zygomaticus) x 4 intensities (high, intermediate, low, neutral) repeated measures ANOVA on the data from the SM condition showed that Sphericity was violated for the intensity levels ($W(5) = .635$, $p < .001$), face muscle sites ($W(9) = .485$, $p < .001$), and the interaction face muscle site*intensities ($W(77) = .025$, $p < .001$), hence, Greenhouse-Geisser adjustment of degrees of freedom was applied. The main effect of *intensity* was significant ($F(2.26,178.18)$

= 7.66, $p < .001$, partial $\eta^2 = .088$, power = .96). Pairwise comparisons between the intensity levels indicated that the EMG activity during processing of neutral faces was significantly lower than during high intensity facial expressions ($p = .001$), during processing of intermediate intensity facial expressions ($p = .002$), and during processing of low intensity facial expressions ($p = .001$). There were no other significant differences (p's > .05); see Figure 49. A figure on the untransformed data can be found in Appendix H.

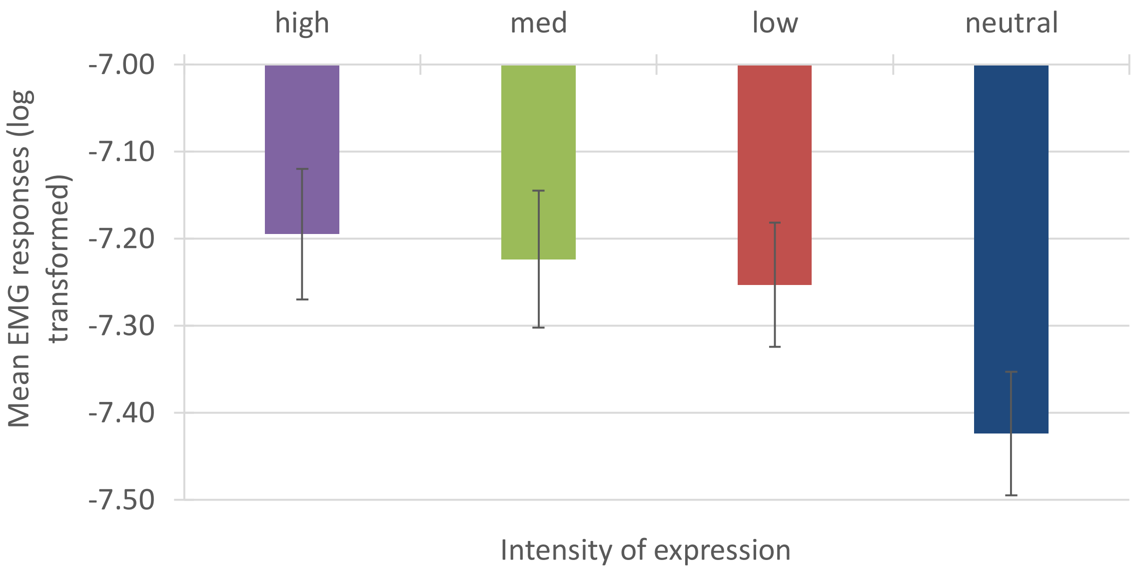


Figure 49. Face EMG responses for each intensity level during spontaneous facial mimicry. Since the data is log transformed, the smaller the negative numbers, the higher the EMG responses. Error bars represent standard errors of the means.

The main effect of *face muscle site* was also significant ($F(2.85,224.99) = 12.60$, $p < .001$, partial $\eta^2 = .138$, power = 1.00). Pairwise comparisons between the face muscle sites showed that overall the EMG responses in the zygomaticus were significantly lower than in the corrugator, depressor, frontalis, and levator (p's < .001). The EMG responses were also significantly higher in the depressor than in the levator ($p = .020$). There were no other significant differences between face muscle sites (p's > .05); see Figure 50. A figure on the untransformed data can be found in Appendix I.

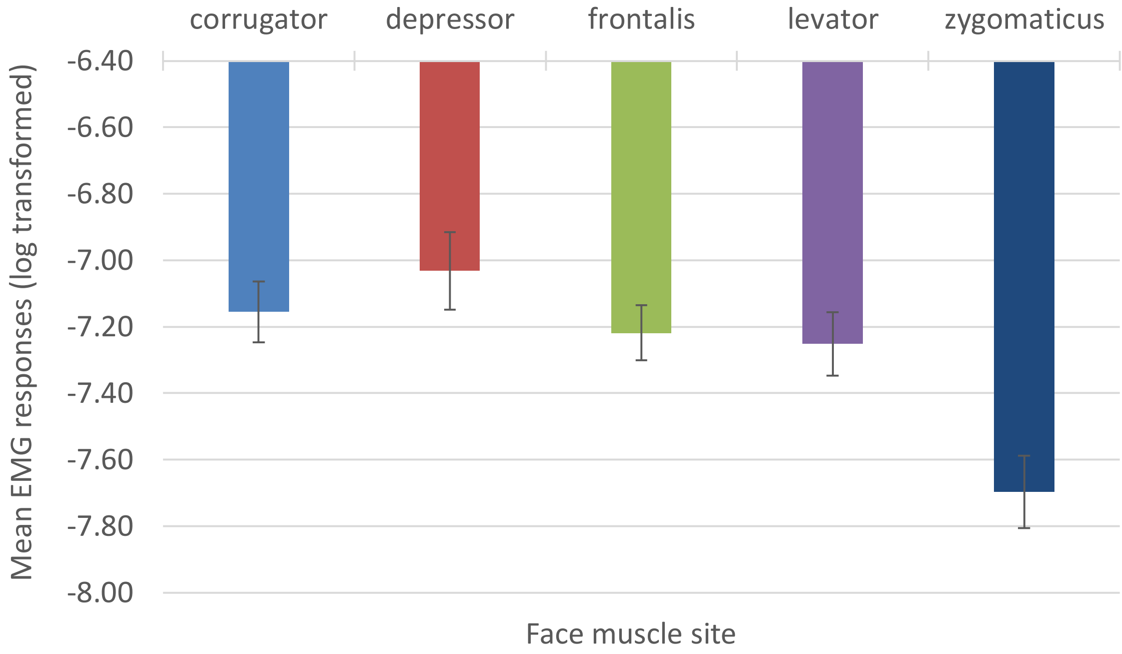


Figure 50. Face EMG responses for each muscle site during spontaneous facial mimicry. Since the data is log transformed, the smaller the negative numbers, the higher the EMG responses. Error bars represent standard errors of the means.

The interaction of *intensity*face muscle site* was significant ($F(7.71,608.77) = 5.71$, $p < .001$, partial $\eta^2 = .067$, power = 1.00). Pairwise comparisons to follow up the significant interaction showed that for the corrugator the activity during high intensity expressions was significantly higher than during the processing of neutral faces ($p = .030$). The activity during low intensity expressions was also significantly higher than during neutral faces ($p = .017$). There were no other significant differences within the corrugator (*p*'s > .05). For the depressor, the activity was significantly higher during processing of high intensity expressions than during neutral ($p = .010$) as well as during intermediate intensity expressions compared to neutral ($p = .018$) and low intensity ($p = .031$). There were no other significant differences within the depressor (*p*'s > .05). For the frontalis, the EMG activity was significantly different between intermediate and low intensity ($p = .016$) and a trend was found for the activity during high compared with low intensity expressions ($p = .068$). There were no other significant differences within the frontalis (*p*'s > .05). For the levator, the EMG activity was significantly higher during processing of high intensity expressions compared to low intensity expressions ($p = .046$) and a trend was found for the difference between the EMG activity during processing of high intensity expressions and neutral faces (p = .062). There were no other significant differences within the levator (p's > .05). For the zygomaticus, the EMG activity was significantly lower during the processing of neutral faces than during the processing of high, intermediate, and low intensity facial expressions (p < .001). A trend was found for higher EMG activity during processing of high intensity expressions compared to low intensity expressions (p = .079). There were no other significant differences within the zygomaticus (p's > .05), see Figure 51. A figure on the untransformed data can be found in Appendix J.

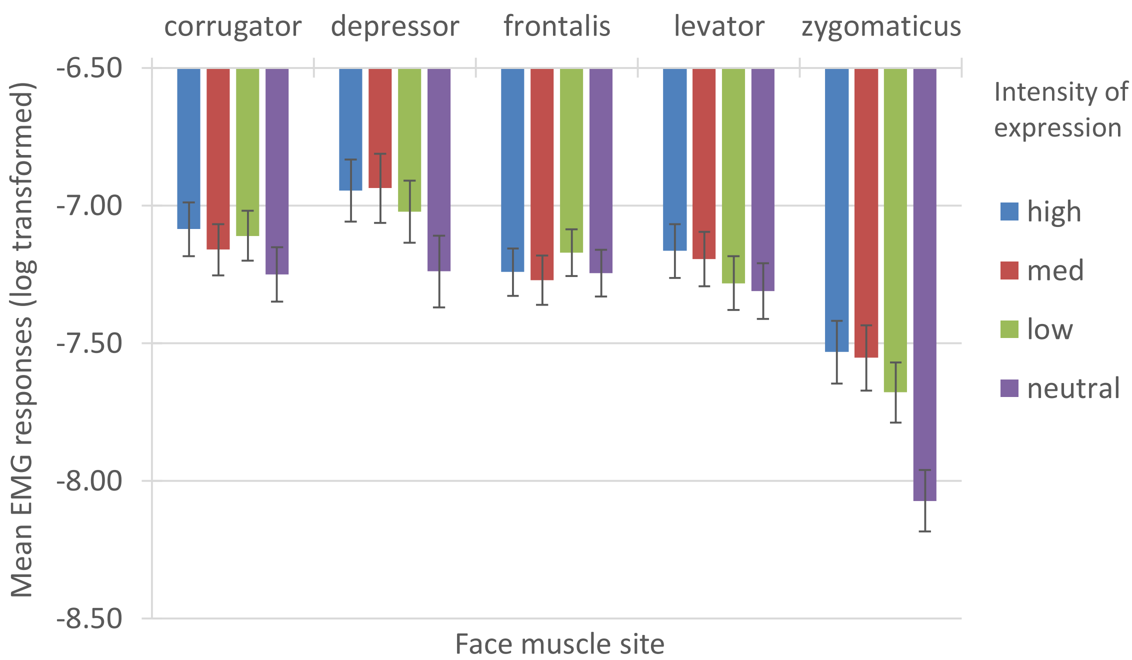


Figure 51. Face EMG responses across intensity levels for each muscle site during the spontaneous facial mimicry condition. Error bars represent standard errors of the means.

Aim 3: Influence of the face movement manipulations on facial emotion recognition.

Influence of the face movement manipulations on accuracy of response.

The results of the repeated measures ANOVAs showed that for most of the emotions the accuracy rates did not differ between the experimental conditions; see Figure 52. That is, for *surprise* the accuracy of response rates obtained within the experimental conditions did not differ significantly from each other ($F(2,166) = .09$, $p = .910$, partial $\eta^2 = .001$, power = .06). For *anger*, the accuracy of response rates also did not differ significantly between the experimental conditions ($F(2,166) = 1.24$, $p = .293$, partial $\eta^2 = .015$, power = .27). For fear, the accuracy of response rates did not differ significantly between the experimental conditions ($F(2,166) = .07$, $p = .931$, partial $\eta^2 = .001$, power = .06). For *sadness,* the accuracy of response rates also did not differ significantly between the experimental conditions ($F(1.83,151.76) = .83$, $p = .429$, partial $\eta^2 = .010$, power = .18); Greenhouse-Geisser adjustment of degrees of freedom was applied due to violation of sphericity assumption. For *surprise* ($F(2,166) = .09$, $p = .910$, partial $\eta^2 = .001$, power = .06), *happiness* ($F(2,166) = .79$, $p = .457$, partial $\eta^2 = .009$, power = .18), *neutral* ($F(2,166) = 1.20$, $p = .304$, partial $\eta^2 = .014$, power = .26), *pride* ($F(2,166) = .47$, $p = .626$, partial $\eta^2 = .006$, power = .13), and *contempt* ($F(2,166) = .56$, $p = .575$, partial $\eta^2 = .007$, power = .14), the accuracy of response rates also did not differ significantly between the experimental conditions. However, a trend was found for *disgust* ($F(2,166) = 2.58$, $p = .079$, partial $\eta^2 = .030$, power = .51). Pairwise comparisons showed that the accuracy of response was significantly higher when participants explicitly imitated the disgust expression ($M = 6.54$, $SD = 2.30$) compared to when facial mimicry was blocked ($M = 6.12$, $SD = 2.25$, p = .038), but not significantly different to the no face movement manipulation condition ($M = 6.26$, $SD = 2.30$, $p = .453$). A trend was also found for *embarrassment* ($F(1.87,154.95) = 2.62$, $p = .080$, partial $\eta^2 = .031$, power = .50). Greenhouse-Geisser adjustment of degrees of freedom was applied due to violation of sphericity assumption. Pairwise comparisons showed that the accuracy of response was significantly higher when participants explicitly imitated the embarrassment expression ($M = 6.01$, $SD = 2.13$) compared to the no face movement manipulation condition ($M = 5.54$, $SD = 1.77$, $p = .021$), but not significantly different to when facial mimicry was blocked ($M = 5.85$, $SD = 2.11$, $p = .484$). The accuracy of response from the blocked mimicry manipulation did not differ significantly from the no face movement manipulation condition (p = .109).

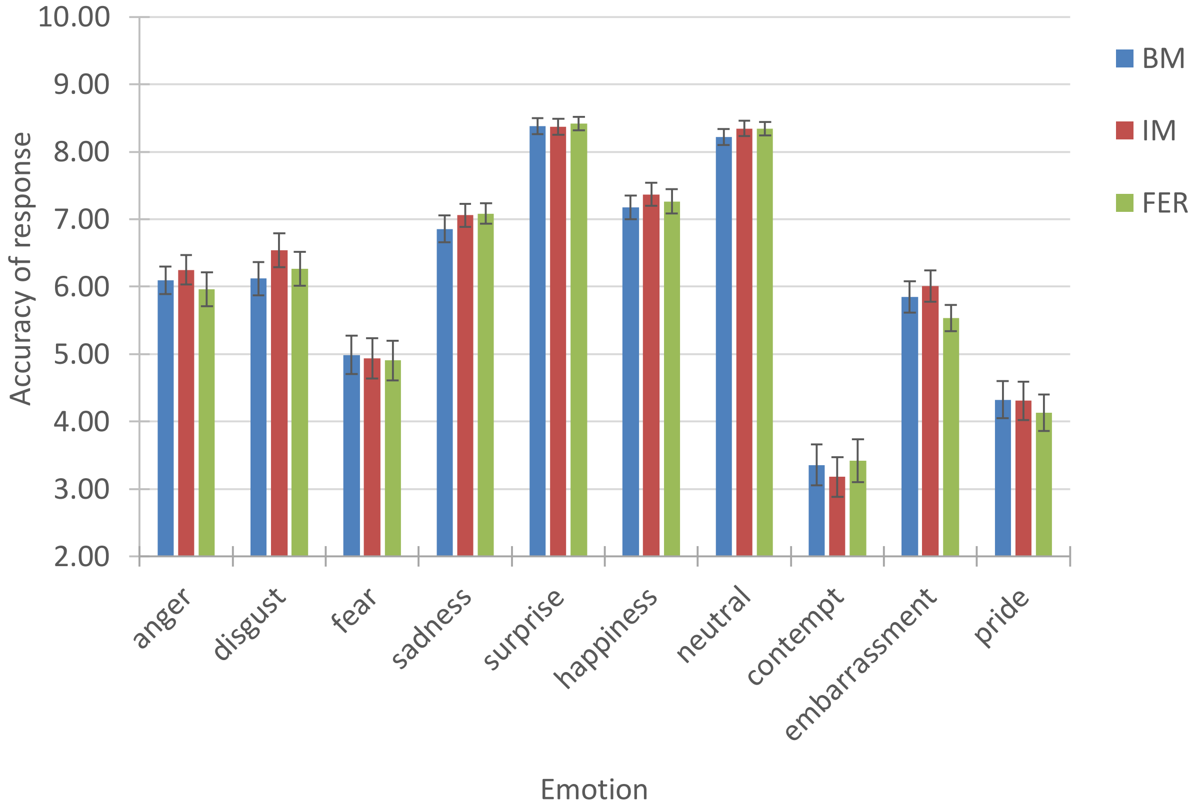


Figure 52. Accuracy of response rates for the emotion categories within each of the three experimental conditions with a facial emotion recognition task (BM = blocked mimicry; IM = explicit imitation; FER = spontaneous facial mimicry with facial emotion recognition). A value of 10 represents 100% accuracy of response. Error bars represent standard errors of the means.

Discussion

The aims of this study were to investigate first whether facial mimicry is emotion-specific with muscle activations similar to when explicitly imitating expressions, second whether the intensity of overserved facial emotional expressions reflects relatively in the intensity of facial mimicry similar to when explicitly imitating expressions, and third to investigate if facial mimicry is a means to facial emotion recognition based on the effect of proprioceptive feedback from face movement manipulations on facial emotion recognition.

Regarding the first aim, spontaneous facial mimicry was indeed observed within this study across five facial muscle sites for the six basic and three complex emotions, showing distinct patterns of face muscle activation for each emotional expression. However, the pattern of EMG responses differed slightly between facial mimicry and explicit imitation in that the order of face muscle site activations from highest to lowest activity was not identical. Nonetheless, the results further support the notion that observed facial emotional expressions are subtly mimicked and in an emotion-specific manner. Regarding aim 2, the varying intensities of facial expression of emotion reflected in the pattern of EMG responses when individuals explicitly imitated the expressions, but not clearly for facial mimicry. It can be concluded that the intensity of the emotion observed does not evidently reflect in the intensity of facial mimicry. Proprioceptive feedback as a result of facial mimicry is thereby less likely to be used for decoding of observed facial expressions of emotion as would be assumed based on the reverse simulation model. To investigate aim 3, the experimental conditions of facial movement manipulation including a facial emotion recognition task were compared based on the achieved accuracy rates by emotion. Trends were found for explicit imitation facilitating recognition of embarrassment and disgust. However, no other facilitations or hindrances due to proprioceptive feedback were found. Facial mimicry did not facilitate facial emotion recognition in comparison to blocked facial mimicry for any of the emotions investigated. It can be concluded that facial mimicry is not a means to facial emotion recognition, again challenging the reverse simulation model. Nonetheless, enhanced proprioceptive feedback from the observed facial emotion compatible muscle activations can selectively facilitate recognition. Additionally, incompatible muscle activations can selectively impair recognition supporting the embodied cognition accounts.

Distinct EMG response patterns during facial mimicry and imitation (Aim 1).

During facial mimicry, distinct patterns of face muscle activations were found for all emotional facial expressions included in the ADFES-BIV, which shows that people subtly mimic not only basic but also complex facial emotional expressions. Precisely, increased EMG responses were found in the corrugator for emotions that include a frown in the facial expression: anger sadness, disgust, and fear. This corrugator activation is in line with the prediction and with the literature as reported in a literature review on facial mimicry (Hess

& Fischer, 2013). Since increased corrugator responses were found for all negative basic emotions, it would seem that corrugator activation is based on valence rather than being emotion-specific as proposed by Hess and Fischer (2013). However, the muscle activations that were found to accompany the corrugator activation for these emotions point towards emotion-specific EMG responses. Results showed that during observation as well as in explicit imitation of surprise faces, activity in the frontalis was increased, which is attributable to pulled up eyebrows and wide-open eyes. This result is in line with the prediction and previous reports (Lundqvist, 1995; Lundqvist & Dimberg, 1995). Against the prediction and reports by Lundqvist (1995), facial mimicry did not emerge in the frontalis during observation of fear faces, although explicit imitation of fear led to increased frontalis activation. That no facial mimicry was found in the frontalis in response to fear faces is however in line with Lundqvist and Dimberg (1995). Unexpectedly, increased zygomaticus activity was found in response to observation of fear expressions, which did not appear during explicit imitation. This zygomaticus activation does not constitute facial mimicry, since facial mimicry requires congruent facial muscle activation and the facial expression of fear does not generally include activation of the zygomaticus. Indeed, zygomaticus activity has only been reported in the literature in association with positive emotions (see Hess & Fischer, 2013). A potential explanation is that some participants might have reacted with a smile to observed expressions of fear due to amusement or to 'regulate' the encoder relating to the social functions of facial expressions discussed in Chapter 1. An alternative explanation is that a facial expression of fear can include the angles of the mouth being pulled outwards. This does not appear as a smile, but the electrodes that were placed above the zygomaticus might have measured some of the electric impulses elicited by that movement due their close proximity. As was discussed in Chapter 3, surface electrodes measure EMG activity of an area rather than one specific spot on the face. In contrast to the passive viewing, during explicit imitation of fear the activity was increased in the frontalis, as would be expected. It can be concluded that a distinct pattern of EMG activation for fear emerged, but more investigations are necessary to learn about the exact muscle activations of facial mimicry in response to fear. Results further demonstrated that the expressions including a smile showed an increase in zygomaticus activity, because it is involved in pulling the corners of the mouth outwards and up. As predicted, zygomaticus activation was hence seen for pride and happiness. This face muscle activation was identical to the pattern that emerged during the explicit imitation condition. It was hypothesised that zygomaticus activity might also result from explicit imitation of and facial mimicry in response to contempt, but this was not found. A potential explanation is that contempt includes a unilateral smile: since recordings were only taken from the left side of the face it is possible that the right side did show such activation. However, since activation of the depressor resulted in response to observing contempt, which also moves the corners of the mouth (downwards), the depressor activity might be indicative of mimicking the mouth movement seen in contempt. During explicit imitation of contempt, the EMG activity was most prominent in the frontalis due to the pulling up of the eyebrow. It was hypothesised that during the observation of expressions including pressing the lips together, i.e. embarrassment and sadness, the activity in the depressor would increase. This activation was found for embarrassment only, whereas during the explicit imitation it was evident in both emotions. Sadness has been claimed to express itself in a loss of muscle tone (Ekman & Friesen, 2003; Oberman et al., 2007), which could explain why during facial mimicry the activity in the depressor did not evidently increase. It was hypothesised that during observation of disgust, EMG responses would be seen in the levator, which was only observed during explicit imitation of disgust. That no facial mimicry in the levator was found in response to disgust was found is in line with the reports by Hess and Blairy (2001), although the finding was reported by Lundqvist and Dimberg (1995). The common denominator between Hess and Blairy (2001) and the current study is that less intense facial expressions were used as stimuli in both investigations. It might be that levator activity is only seen in response to high intensity expressions of disgust. Not many have investigated facial mimicry in the levator in response to disgust. More of such investigations are needed to draw a conclusion about whether facial mimicry in response to disgust occurs in the levator and at which level of intensity. Overall, these results suggest that specific face muscle activations seem to occur in response to the emotion categories included in the ADFES-BIV. Even though the face muscle activations differed slightly during explicit imitation and facial mimicry, the occurrence of differentiated EMG response patterns suggests that facial mimicry is indeed emotion-specific. These results support the view of emotions as distinct entities as proposed by Ekman (1992a, 1992b) and as including distinct physiological response patterns.

Since facial mimicry occurred in response to the facial emotional expressions and no instruction was given other than to watch the videos, the occurrence of facial mimicry could be explained by the activity of mirror neurons. Molenberghs et al. (2012) reported that tasks including stimuli showing emotional facial expressions increase brain activity in the regions associated with emotional processing (i.e. amygdala, insula, and cingulate gyrus). The observation of a specific emotional facial expression has been found to lead to brain activation in the same brain regions of the observer (e.g. Enticott, Johnston, Herring, Hoy, & Fitzgerald, 2008; Lee, Josephs, Dolan, & Critchley, 2006; Leslie, Johnson-Frey, & Grafton, 2004; Montgomery & Haxby, 2008; van der Gaag, Minderaa, & Keysers, 2007). This evidence suggests that observing facial emotional expressions and executing them activate the same brain areas: mirror neurons might be the underlying cause. This is further evidenced by Schilbach, Eickhoff, Mojzisch, and Vogeley (2008) who measured spontaneous facial mimicry via EMG in one experiment, and brain activation whilst watching facial expressions in another experiment. When comparing the results of both experiments, they found increased brain activity in areas of motor responses, as well as brain areas associated with social cognition at the same time as facial mimicry occurred. Likowski et al. (2012) combined measuring facial mimicry via EMG and brain imaging in a single experiment. They replicated the findings of mirror neuron activity in areas associated with perception and execution of facial expressions. They further found facial mimicry to be associated with areas involved in emotion processing. Those results suggest that facial mimicry is based on mirror neuron activity.

Intensity of observed expression and intensity of EMG activity during explicit imitation and facial mimicry (Aim 2).

Since it was assumed that facial mimicry is the result of mirror neuron activity, it was investigated with the current study whether the intensity of observed facial expressions reflects in the EMG responses during facial expression production. The expected pattern of EMG responses was found for the explicit imitation condition, but not for the facial mimicry condition. During explicit imitation of the observed facial expressions, the highest EMG responses occurred in response to high intensity facial expressions, followed by intermediate intensity facial expressions, and then low intensity facial expressions, with the lowest EMG activity during explicit imitation of neutral faces. This pattern is in line with the hypothesis and demonstrates that individuals imitated not only facial features (as discussed above), but also the intensity of the observed facial expressions. In contrast, whereas the EMG responses during facial mimicry were significantly higher when observing high intensity facial expressions compared with neutral facial expressions, the increments between the intensity levels were less pronounced for facial mimicry than when explicitly imitating the facial expressions. In fact, no significant differences were found in EMG responses for facial mimicry of low, intermediate, and high intensity facial expressions of emotion. One potential explanation for these results is that maybe the intensity levels of the ADFES-BIV were not different enough from each other in regards to intensity of expression for the generally subtle response of facial mimicry to lead to significant differences between the intensity levels. Indeed, visual inspection of the graph displaying the EMG responses for facial mimicry shows slight differences in activation between the intensity levels in the predicted direction, which however failed to reach significance. Future research should seek to re-investigate this by employing stimuli that are more distinct from each other in regards to intensity of facial expression displayed. However, it is possible that the facial mimicry response only includes facial features and not the relative intensity of facial expressions. Nonetheless, with facial mimicry assumed to have its roots in the activity of mirror neurons within the mirror neuron system, it could be expected that the same pattern that was evident during explicit imitation would be present during the facial mimicry period if mirror neurons literally mirror what is observed. The results from the current study lead to the assumption that facial mimicry could be the result of a simulation process based on the observer's prior knowledge and experience as suggested by embodied cognition accounts (Niedenthal, 2007) rather than an exact mirroring. There are representations, concepts of objects, actions, emotions etc. in our brains, which are formed or acquired early in life. Sensory experiences from all modalities (motor, sensory, and affective) are memorised in these representations. Any subsequent exposure to or the memory of such concepts activates the respective representations. That is, when faced with those objects, actions, or emotions, the representation is triggered and the state when the knowledge was initially acquired re-lived. It is possible that visual sensory input following the observation of an emotional facial expression is first classified as emotional information. As a result, the emotion representation is elicited including all characteristics (from changes in the autonomous nervous system to facial expression) by activating motor areas in the brain that are associated with the observed emotional expression. This is comparable to Ekman's (1992) understanding of emotions and what he called 'affect programme'. As a result of the elicited affect programme, motor areas in the brain might be activated resulting in facial mimicry, independent of the intensity of observed facial emotional expression.

Influence of proprioceptive feedback on facial emotion recognition (Aim 3).

The current study investigated what effect face movement manipulations and the resulting proprioceptive feedback would have on accuracy of response within facial emotion recognition. It was hypothesised that explicitly imitating facial expressions would facilitate recognition more than passive viewing where spontaneous facial mimicry occurs based on the enhanced proprioceptive feedback. It was further expected that the accuracy rates would be higher during the no face movement manipulation than in the blocked facial mimicry condition for the emotions where the mouth area carries the most emotional cues. The results were rather limited. However, disgust faces were better recognised when they were explicitly imitated than when facial mimicry was blocked, but no difference in accuracy of response was found compared to the no face movement manipulation condition. In contrast, previous studies found disgust recognition impaired when mouth movements were manipulated compared to no face movement manipulations (Oberman et al., 2007; Ponari et al., 2012). It has to be noted that disgust should be the emotion affected the most by holding a pen in the mouth, since this action greatly activates the depressor. The depressor is the antagonist to the levator and therefore an activated depressor blocks muscular movements of the levator (the main muscle included in the expression of disgust as it wrinkles the nose). As a result, it was difficult for participants to activate the levator whilst the depressor was activated. With this particular method of blocking facial mimicry, it can therefore be expected that recognition of disgust would be affected as it prevents nose wrinkling. It can be concluded that when muscular feedback is completely blocked, recognition is diminished. This result implies that facial mimicry was necessary for facial emotion recognition and would support the reverse simulation model. However, facial mimicry had no facilitating effect on facial emotion recognition compared to blocked facial mimicry condition in the current study. Thus, it is more plausible that the with the observed emotional expression incompatible muscle activation interferes with recognition based on the contradictory enhanced proprioceptive feedback. The current study showed that only enhanced proprioceptive feedback from face muscles influenced facial emotion recognition and not subtle proprioceptive feedback as from facial mimicry.

The current study found higher accuracy rates for embarrassment when participants explicitly imitated the expression as opposed to simply observing the expression and thereby allowing for spontaneous facial mimicry. It is possible that explicit imitation of the facial expression of embarrassment and the resulting enhanced proprioceptive muscle feedback makes the expression less ambiguous, which helps recognition. However, blocking facial mimicry did not diminish the accuracy rates for embarrassment. This might be due to the fact that the depressor was highly activated during blocking and therefore maybe even helped to identify embarrassment, as embarrassment was found within the current study as the emotion where depressor responses increased most during spontaneous facial mimicry compared to the other muscles. Since for no other emotion category any effects of face movement manipulations on accuracy rates were found, it seems even more so that facial mimicry is not a component of the facial emotion recognition process, in line with the suggestion by Hess & Fischer (2013, 2014). This conclusion fits the claim that facial mimicry is a by-product rather than a necessity in the recognition process (Rives Bogart & Matsumoto, 2010). Facial mimicry of emotional facial expressions could be a by-product resulting from the activation of an emotion representation elicited by observing an emotional facial expression, as explained above in the section on facial mimicry in relation to intensity of observed facial expressions. When an emotion representation is elicited, facial mimicry would be the result of an activated (motor) representation or affect programme of the emotion and therefore constitute more of a by-product of emotion processing instead a necessity for facial emotion recognition. Nonetheless, it has to be noted that holding a pen in the mouth in a kiss-like position in the current study induced the most activity in the depressor, also affected the levator and zygomaticus, but this manipulation left the muscles of the upper face unaffected. This partial blocking of facial mimicry can therefore only have limited effect on the decoding of emotions. Ideally, all facial muscles would have been blocked, but this proves difficult in practice. Overall, the current study did not find support for the reverse simulation model, but found evidence for body-cognition-interactions as proposed by embodied cognition accounts in that enhanced proprioceptive feedback from the face muscles can slightly influence emotion judgements.

Limitations.

The current study had several strengths, like applying dynamic stimuli of varying intensity of facial expression and a wider range of emotions, the advantages of which have been discussed in previous chapters. However, by including varying levels of intensity of facial expression in the task, but presenting each stimulus once, only three trials were available per intensity level for each emotion in each experimental condition. This limited the analysis regarding the influence of observed intensity of facial expression on the intensity of facial mimicry. This was the reason why the data was collated over emotions within each intensity level. Whereas this approach worked for the data on explicit imitation, some effects might have been averaged out for the facial mimicry data where responses were much smaller in general. Instead of having 90 trials per experimental condition, it would have been beneficial to have at least 180 trials (presenting the stimuli twice) or to apply a between-subject design. Another problem that arose from the small number of trials per intensity level for each emotion was that normal EMG values were flagged as extreme values for the emotion categories. It seems more appropriate to apply a between-subject design to increase the number of trials per intensity level, so that the distribution shifts a little and the outlier fall within the normal range. Within the current study, standardisation was applied at category level. Whereas it would have been appropriate to standardise each trial, the just explained issue led to very high *z*-values, which would have to be excluded. It was therefore decided to standardise the emotion categories made of nine trials and not the individual trials themselves.

That a within-subject design with four experimental conditions was applied within the current study was thought to be a strength of the study. The advantage of a within-subject design is that the found effects are the result of the experimental manipulations and not due to potential differences between samples as can be the case in between-subject designs, thereby reducing the error variance. However, the instruction to explicitly imitate the observed facial expressions turned out to have a lasting effect on participants. Some participants showed the same pattern of activation in the BM and FER condition than during the IM condition. This indicates that explicit imitation was carried out in the other manipulations as well. Some participants reported after the experiment that they continued imitating throughout the following manipulations (participants 26, 41, 42, 45, 74, and 86), which only occurred in versions of the experiment where the IM condition was not the last experimental condition (version 4, version 5, version 6, version 2). This might have affected some of the results within the versions of the experiment where IM was not the last experimental condition. This is important to note and means that the IM condition should always be last, although that is problematic regarding the accuracy rates that are generally higher towards the end of the experiment (see Appendix F). Due to the increase in accuracy rates over the course of the experiment, no participant could be eliminated. This is because elimination of participants would have led to uneven numbers of participants for each version of the experiment with more participants in the versions where IM was the last experimental condition; the improvement effect would have biased the results of the experimental manipulations. Precisely, the effect of explicit imitation would be inflated, as accuracy rates are generally higher towards the end of the experiment.

Another strength of the current study is that the sample size was fairly large contributing to power of the statistical analyses and reliable results. However, despite the attempt to block as many muscles in the face as possible from engaging in facial mimicry, the upper face was unaffected by the blocking. As mentioned earlier, ideally, muscular activity that is incompatible with any emotional facial expression would have been induced in the whole face to really investigate the effects of 'blocked' facial mimicry on facial emotion recognition. However, there seems to be no applicable option to block the muscles of the upper face. In some investigations, participants were instructed to suppress any facial reactions (e.g. Schneider et al., 2013), but this constitutes a rather unsystematic and unreliable approach. Researcher have also attempted to block upper face muscles by instructing participants to perform certain facial movements (e.g. Ponari et al., 2012). First of all, it is still more likely that any performed action (e.g. drawing eye brows together) resembles an emotional facial expression even if only partially and can therefore not considered an option to block facial mimicry. Secondly, instructing participants to perform a facial action requires a constant mental effort to keep the action executed. It is likely that the activity would therefore vary greatly. This is not to say that the pen holding did not require any mental effort, nor that the activity did not vary. However, the pen was visible and acted as a constant reminder. The only true option to blocking facial mimicry would be injection of Botox, which is difficult from ethical point of view.

The next chapter provides a summary of all studies conducted and presented within this dissertation next to discussion of the results across studies. Results are discussed within relevant theories that have been presented in previous chapters and limitations of the conducted research and directions for future research are provided.

GENERAL DISCUSSION

CHAPTER 10—Summary and discussion from studies of dissertation

The overall purpose of this dissertation was to investigate emotion processing from video stimuli of facial emotion portrayed at varying intensity levels of facial emotional expression. Therefore, the ADFES-BIV, a video stimulus set of facial emotion including varying levels of expression intensity was developed (Chapter 4) and validated (Chapter 5-7). The current research project also aimed to assess differences in individuals' ability to recognise emotions from videos of faces displayed at varying levels of intensity. Therefore, the ADFES-BIV was applied to assess variations in facial emotion recognition ability in the general population based on sex (Chapter 6) and in a clinical population comparing individuals with high-functioning ASD to controls (Chapter 8). Another aim was to gain knowledge about facial mimicry in response to videos of facial emotional expressions of varying intensities and to test the reverse simulation model's assumption that facial mimicry is a necessity within the facial emotion recognition process. The aim was explored by investigating the effect of face movement manipulations on facial emotion recognition (Chapter 9). This chapter will first briefly review the findings from each study chapter and then discuss the findings across experimental investigations.

Summary of results from studies of dissertation

Chapter 4.

Pilot study 1 from Chapter 4 aimed to select a video stimulus set, on the basis of which video stimuli of varying intensities of expression were created, and to evaluate these newly created videos regarding recognisability on individual item level. Based on the ADFES (van der Schalk et al., 2011), three intensities of facial expression were created (low, intermediate, and high). It was expected that the recognition rates of all 360 videos would be greater than a chance level of responding of 10%. The results of the initial evaluation on a student sample showed that 14 of the 360 videos failed to be recognised above the chance level of responding. The 14 videos were re-edited with the aim of increasing their recognisability. Pilot study 2 from Chapter 4 aimed to test if individuals with high-functioning autism would be able to recognise the emotions from the videos, measured on their accuracy of response rates for the emotion categories included in the stimulus set. (Note that pilot study 2 included the initial set of videos, not the re-edited videos, since there was an overlap between the availability of ASD participants and the re-editing process; participants came from the Autism Summer School which was held at the University of Bath). It was expected that the recognition rates would be greater than the chance level of responding for the categories included in the facial emotion recognition task. This feasibility test was necessary since individuals with ASD often show deficits in facial emotion recognition (e.g. Harms et al., 2010; Uljarevich & Hamilton, 2012). Those deficits combined with an increased task difficulty due to the inclusion of subtle facial expressions could have led to performance at chance level of responding for the low intensity expressions. Such an outcome would have rendered the stimuli unsuitable for assessment of facial emotion recognition in ASD. The results from the ASD pilot study showed that the individuals with ASD had recognition rates for the emotion categories included in the stimulus set that were higher than the chance level of responding. Based on the results from the pilot studies in Chapter 4 it was decided to validate the videos on a larger sample from the general population.

Chapter 5.

The study presented in Chapter 5 aimed to validate the intensity levels of the ADFES-BIV and to confirm the intensity levels as distinct categories. This was tested based on participants' judgements of perceived intensity of the facial expressions in the videos on a student sample. It was expected that the high intensity facial expressions would be judged as more intense than the intermediate intensity expressions and that the intermediate intensity facial expressions would be judged as more intense than the low intensity expressions. This pattern was expected to emerge also within the emotion categories. The pattern of low intensity expressions being judged as less intense than intermediate intensity expressions and the being in turn judged as less intense than high intensity expressions was not only true for the intensity levels overall but also at the level of individual emotions. Neutral faces were perceived as the least intense. Results demonstrated that the three intensity levels were indeed perceived as distinct based on subjective judgements of intensity of expression.

Chapter 6.

The aim of the study presented in Chapter 6 was to validate the ADFES-BIV (including the 14 re-edited videos), which included testing recognition rates of the different emotion and intensity categories against the 10% chance level of responding and comparing emotion and intensity categories between themselves. Accuracy of response as well as response latencies were examined to validate the intensity levels. A further aim was to test the ADFES-BIV's sensitivity to reveal group differences in facial emotion recognition in the general population by comparing recognition rates and response latencies of males versus females on all its categories. It was expected that all emotion and intensity categories would be recognised significantly better than the chance level of responding. For the intensity levels it was hypothesised that the low intensity facial expressions would have lower recognition rates and longer response times than the intermediate intensity expressions, which in turn would have lower recognition rates and longer response times than the high intensity expressions. A female advantage over males based on accuracy rates and response latencies in facial emotion recognition ability was expected. The results of the validation study from Chapter 6 showed that the emotion categories overall and within each intensity level of the ADFES-BIV were recognised significantly above the chance level of responding. For the three intensity levels of expression, the results revealed a pattern of greater accuracy rates for the high intensity expressions compared to the intermediate intensity expressions, which in turn had higher accuracy rates compared to the low intensity expressions. Response latencies followed the same pattern with fastest responses to high intensity expressions and slowest to low intensity expressions. A female advantage in facial emotion recognition was replicated. Precisely, females achieved higher accuracy rates and responded faster than males. Together, these results indicate an effective validation of the ADFES-BIV as well as suitability of the ADFES-BIV to investigate group differences in facial emotion recognition ability in the general population.

Chapter 7.

Since the videos of the ADFES-BIV have varying exposure times between intensities, a study was carried out to test the effect of exposure time of the emotional facial expression on accuracy rates and response latencies of the intensity levels of the ADFES-BIV. It was expected that the results from the study presented in Chapter 7 would replicate the pattern from the study in Chapter 6, with highest accuracy rates and fastest responses to the high intensity videos, and lowest accuracy rates and slowest responses to the low intensity videos. The results from the study in Chapter 7 showed the same pattern as with the ADFES-BIV for the accuracy rates, namely greater accuracy rates for the high intensity expressions compared to the intermediate intensity expressions, which in turn had higher accuracy rates compared to the low intensity expressions. The same was true for the comparisons of response latencies between the intensity levels, although the difference in response times between low and intermediate intensity expressions was not significant. These findings indicate that exposure time has little, if any, effect on accuracy of response or response latencies. The study from Chapter 7 demonstrated that the intensity levels of the ADFES-BIV are distinct categories and that the differences in accuracy of response between the intensity levels are the result of the differences in intensity of expression rather than exposure time variations.

Chapter 8.

The aim of the study presented in Chapter 8 was to investigate facial emotion recognition in ASD by comparing facial emotion recognition in high-functioning ASD to controls, by comparing the confusions between emotion categories made by the ASD group statistically to those made by typical individuals, and by investigating the confusions made by the ASD group in regards to sensitivity and specificity in facial emotion recognition. It was hypothesised that individuals with ASD would show an overall deficit in facial emotion recognition compared to controls, that the deficit in facial emotion recognition would also be specific to certain emotions (negative emotions) and not others (positive emotions), and that the recognition of the emotions would be influenced differently by the level of intensity in the two groups. It was further expected that the ASD group would show significantly greater impairments in specificity based on the weak central coherence theory (Frith, 1989/2003), i.e. be more likely than typical individuals to confuse emotions that overlap featurally in their facial expression (e.g. anger and disgust, fear and surprise), and that individuals with ASD would show greater impairments in sensitivity, i.e. would have problems with detecting emotions from faces evidenced by significant confusions with the category neutral. In line with these expectations, an overall facial emotion recognition deficit in individuals with ASD compared to controls based on accuracy rates was found. Against the prediction, the facial emotion recognition deficit in ASD did not differ significantly from controls for the individual emotion categories or intensity levels of expression tested. As expected, the intensity of facial emotional expressions was differently influenced by intensity in individuals with ASD compared to controls. Fear and sadness were the only emotions identified where the ASD group achieved significantly lower accuracy rates on all three levels of intensity. For anger and embarrassment, the ASD group only performed worse than controls at facial emotion recognition at the lower intensities; for no other comparison of the group means statistical significance was reached. In line with the prediction, individuals with ASD showed deficits in specificity and sensitivity in facial emotion recognition. The ASD group confused fear at high intensity with surprise significantly more often than the controls. Individuals with ASD perceived disgust, sadness, and anger as neutral to a significantly greater extent than the controls when expressed at low intensity. Together, the results show that individuals with ASD have a deficit in facial emotion recognition that is the result of difficulties in perceiving as well as discriminating emotional content from faces.

Chapter 9.

One aim of the study presented in Chapter 9 was to test whether a distinct pattern of EMG activation across the five face muscle sites for facial mimicry and explicit imitation occurs for each of the six basic and three complex emotions included in the ADFES-BIV by comparing the EMG responses of the five face muscle sites to each other for each emotion. Another aim was to investigate whether the intensity of observed facial expression of emotion reflects relatively in the intensity of facial mimicry and explicit imitation in participants by comparing the EMG responses to the intensity levels of the ADFES-BIV to each other. A further aim was to test if facial mimicry is a necessity of facial emotion recognition as suggested by the reverse simulation model, by examining the effects of proprioceptive feedback (from facial mimicry, explicit imitation, and blocked facial movements) on facial emotion recognition of specific emotions. For this purpose, the accuracy rates for the experimental conditions were compared for each emotion. It was expected to find distinct patterns of face EMG responses to each of the emotion categories of the ADFES-BIV for facial mimicry and explicit imitation of observed emotional expressions. It was also expected that greater intensity of observed facial expressions would be evident with higher EMG responses during explicit imitating of facial expressions and for facial mimicry. Based on the reverse simulation model and the assumed influence of proprioceptive face muscle feedback on recognition, it was further anticipated that the lowest accuracy rates from the blocked mimicry condition and the highest accuracy rates would emerge from the explicit imitation condition with the no face movement manipulation condition in between. In line with the prediction, for each of the nine emotions patterns of face muscle activation were found that were distinct from the other emotions and in synchronisation with the emotional facial expressions observed. That is, facial mimicry occurred in response to observed facial expressions of emotion for all emotions investigated. These results are in line with distinct emotion theories (e.g. Darwin, 1872/1965; Ekman, 1992a, 1992b; Izard, 1977). Against the prediction, the intensity of facial mimicry did not significantly increase from intensity level to intensity level. In contrast, during explicit imitation the increase in EMG responses from intensity level to intensity level of observed facial expression was significant. That is, the intensity of observed facial feature activations is not included in the simulation of observed facial emotional expressions. Against the expectation, facial mimicry did not have a facilitating effect on facial emotion recognition for any of the emotions tested compared to the blocked mimicry condition. This result challenges the reverse simulation model. However, enhanced proprioceptive facial feedback from face movement manipulations can facilitate and hamper facial emotion recognition. This conclusion is based on the findings that disgust recognition was improved when participants explicitly imitated the expression compared to when facial movements were blocked, and that recognition of embarrassment was improved during explicit imitation compared to the no face movement manipulation condition. Together, the results show that facial features of observed facial expressions of emotion are mimicked, but not the relative intensity of facial feature activation, and that facial mimicry is not a necessity for facial emotion recognition.

Discussion of results from studies of dissertation

Recognising facial emotional expressions from videos with varying intensity.

The results from the studies applying the ADFES-BIV to assess facial emotion recognition strongly suggest that accuracy rates are influenced by the intensity of facial expression. This is evidenced by the higher recognition rates reported by the developers of the ADFES videos (van der Schalk et al., 2011), which always displayed apex facial emotional expressions, than the recognition rates from the ADFES-BIV. The authors reported an overall recognition rate of 85% for the Northern European ADFES videos and recognition rates ranging from 69% to 95% for the emotion categories included. The accuracy rates from the ADFES-BIV (here on the example of the results from the validation study in Chapter 6) were lower overall (69%) and for the emotion categories with a range of 34-92%, since varying levels of intensity of expression were included. Facial emotional expressions of lower intensity are harder to recognise and therefore result in lower recognition rates, which lower the accuracy rates overall and for the emotion categories. All emotion categories of the ADFES-BIV were affected by intensity, although the magnitude of the effect varied between emotion categories. Intensity of expression seems to be an influencing factor on the recognisability of a facial emotional expression.

A female advantage was found in the study from Chapter 6 across the three intensity levels of the ADFES-BIV, in contrast to previous research based on morphed static images of varying expression intensity where the female advantage was only present for subtle facial expressions (Hoffman et al., 2010). Hoffman et al. (2010) suggested that the female advantage only becomes apparent when task difficulty is high, which is particularly true for subtle expressions. It has to be noted that Hoffmann et al. (2010) defined the intensity levels differently. Precisely, Hoffmann et al. (2010) defined the low intensity category as including facial emotional expressions of 40-50% intensity, the medium intensity category as 60-70%, and the high intensity category as 80-100%. By contrast, the subjective judgements of the intensity levels for the ADFES-BIV from the study presented in Chapter 5 yielded on average 42% for the low intensity level, 55% for intermediate, and 66% for high intensity level. That is, the high intensity level of the ADFES-BIV is equivalent to the medium intensity in Hoffmann et al. (2010) and the intermediate intensity level of the ADFES-BIV is equivalent to the low intensity in Hoffmann et al. (2010). (It has to be noted though that the subjective intensity judgements of the ADFES-BIV do not necessarily have to align with the intensity levels that were created by using the morphing technique, which is mathematically based). If the ADFES-BIV includes lower expression intensity overall, then task difficulty was increased with the ADFES-BIV. Given that slightly more difficult tasks are more likely to reveal potential group differences, this explains the female advantage over males found in Chapter 6 across all intensity levels. The ADFES-BIV has proven to be a sensitive measure for assessing group differences in facial emotion recognition by comparing males to females on facial emotion recognition ability.

The accuracy of response rates found for the intensity levels of the ADFES-BIV were consistent across the studies within this dissertation and comparable to the accuracy rates reported from other video stimulus sets including intensity variations. The accuracy rates that were achieved with the ADFES-BIV were of similar magnitude across the studies from Chapter 5 and 7 (low = 56-57%, intermediate = 71-73%, high = 80-82%). The DaFEx (Battocchi et al., 2005) is an example video stimulus set including three intensity levels of expression for the six basic emotions. The reported recognition rates for the intensity levels were 67% for the low intensity expressions, 77% for the medium intensity expressions, and 83% for the high intensity expressions. When looking at the accuracy rates for the intensity levels of the ADFES-BIV including both, basic and complex emotions, the accuracy rates of the ADFES-BIV are slightly lower than from the DaFEx. However, the accuracy rates for the intensity levels of the ADFES-BIV align with the DaFEx when only looking at the six basic emotions (ADFES-BIV: low = 67%, intermediate = 79%, high = 84%). This finding demonstrates that including complex emotions in a face emotion stimulus set lowers the recognition rates it yields. Together, the accuracy rates from the intensity levels of the ADFES-BIV show that the accuracy rates are comparable to other video stimulus sets of facial emotion at varying levels of intensity.

An explanation for the influence of expression intensity on accuracy rates is that less emotional information is available in expressions of lower intensity than ones of high intensity. That is, the more emotional information is available, the easier the emotion is to recognise. Facial expressions of high intensity carry more emotional information, since facial features and configurations are more apparent than in low intensity facial expressions of emotion. Since decoding of facial emotional expressions includes analysis of observed facial features and configurations, more apparent features and configurations facilitate recognition. Based on morphed images, Hess et al. (1997) found that with increasing intensity of a facial emotional expression, the recognisability of the facial expression also increases. This was also shown with the ADFES-BIV from the studies presented within the current dissertation, as throughout the studies low intensity facial expressions led to lower accuracy rates than high intensity facial expressions, with the accuracy rates from the intermediate intensity facial expressions falling in between. The differences found in recognition rates between intensity levels of expression can be explained with the amount of emotional information that is available in the facial expressions.

A potential theoretical explanation for differences in recognition rates between intensity levels of expression is provided by the simulation theory, particularly the reverse simulation model. Observed facial expressions are simulated with brain activation in the observer equivalent to the brain activation in the encoder. Facial mimicry is the result of this simulation. If this simulation process includes not only facial features, but also the relative intensity of their activation, a potential explanation for the better recognisability of facial expressions that are expressed at high intensity, is that more intense facial mimicry is underlying. More intense facial mimicry in turn results in more proprioceptive feedback from the face muscles to the brain, information that is then utilised for decoding the observed facial emotional expression. However, the results from the study presented in Chapter 8 do not support this assumption, since the EMG responses from the face muscles investigated were not significantly different between the intensity levels of expression during the assessment of facial mimicry. That is, the intensity of facial mimicry does not vary with the intensity of observed facial expressions. It can be concluded that it is the amount of emotional information available in the observed face that results in lower accuracy rates to low intensity facial emotional expressions than high intensity expressions and not the intensity of facial mimicry in an individual's own face. If the intensity of observed facial emotional expressions affects recognition, but leads to similar brain activation in motor areas of the brain (inferred from the results on facial mimicry), then it seems that the facial feedback from facial mimicry is not the mechanism used for facial emotion recognition. The arising question is what mechanism is underlying the lower accuracy rates to subtler facial expressions compared to high intensity expressions. It would be interesting to investigate the brain activity during processing of facial emotional expressions with varying expression intensity. It is still possible that more intense facial emotional expressions lead to higher brain activity in the visual cortex and maybe even in the limbic system (associated with emotion processing), but not in motor areas that cause the facial mimicry response. If it is possible, investigation of the time sequence of brain activation in the various brain regions would provide useful information on this topic. The reverse simulation model suggests that facial mimicry precedes the emotion judgement, but it is also possible that facial mimicry is the result of an emotion judgement. If motor activity is last in the sequence of processing emotional facial expressions, then that would suggest observed expression intensity only influences what is perceived (e.g. is a subtle smile perceived as such or as a non-emotional mouth movement?). Based on the perceptual input, the judgement of the emotion might take place and after the judgement the respective emotion representation gets elicited leading to facial mimicry. With facial mimicry being the result of a simulation based on an emotion representation within the observer and their experiences, the independence of the intensity of facial mimicry from the observed intensity of expression would be explained. This scenario would also imply that facial mimicry does not have to match the observed facial expression. Future research should compare the pattern of facial mimicry from correct trials to incorrect trials of facial emotion recognition to shed light on this.

Instead of explaining the differences in accuracy of response between the intensity levels of the ADFES-BIV with intensity of observed facial emotional expression, the differences in accuracy of response can be explained by exposure time to emotional content, which varied between the intensity levels. It is possible that longer exposure time to emotional content in the high intensity facial emotional expressions is underlying the higher accuracy rates from the high intensity facial expressions than from the lower intensities. However, the results from the study presented in Chapter 7 showed that this is unlikely. The exposure time to emotional content and neutral facial expressions was kept constant with the first-last frame stimuli and the results still showed that accuracy rates were lower and response latencies longer to low intensity facial expressions than high intensity facial expressions. Bould et al. (2008) also found that exposure time has less of an influence on facial emotion recognition. Even when exposure time to emotional content is controlled for, the intensity of facial expression seems to modulate the expression's recognition. However, the progression of the facial expressions was not visible in the first-last frame stimuli, as all frames but the first and last of the ADFES-BIV videos were omitted. The first-last frame stimuli include motion in that the facial expression changes from neutral to the end point of the emotional expression. The neutral face serves as a point of reference and the perception of change can be used for decoding of the facial emotional expression. In contrast, the ADFES-BIV videos do not only provide the perception of change but also show the unfolding of the facial expressions. As a consequence, emotional information was omitted in the first-last frame stimuli that might otherwise would have benefited recognition. This is in line with the finding that dynamic stimuli of facial emotional expression generally lead to higher recognition rates than static images of facial emotion, which has been shown experimentally in several studies comparing accuracy rates from facial emotion recognition tasks including static images to the accuracy rates from dynamic images (literature review by Krumhuber, Kappas, & Manstead, 2013). A similar observation was made from the accuracy rates from the first-last frame stimuli (study in Chapter 7) compared to the accuracy rates of the ADFES-BIV validation (study in Chapter 6). The accuracy rates for the intensities from the first-last frame stimuli were consistently lower and response times were consistently longer than from the ADFES-BIV videos. (Note that this is not based on statistical comparisons, but the samples for both studies were drawn from the same population, which adds some validity to this observation). It seems that the unfolding of facial emotional expressions in videos provides further information exceeding the information from the perception of change that is used during decoding and facilitates recognition, accuracy rates and response times. Bould et al. (2008) investigated this experimentally by manipulating temporal characteristics of the six basic emotions and examining the effects on accuracy rates. Since the manipulations of speeding up and slowing down the sequences of facial emotional expressions decreased the accuracy rates, Bould et al. (2008) concluded that temporal characteristics of an unfolding emotion provide information about the emotion (see also Kamachi et al., 2001). It can be concluded that emotions have specific temporal characteristics that are considered during decoding of facial emotional expressions. These temporal characteristics are only provided by video stimuli of facial emotion, not by static images.

The inclusion of motion in stimuli of facial emotion by presenting unfolding facial expressions not only provides emotion-specific temporal characteristics, but also seems to facilitate the visual processing of the emotional expression per se. For example, Yoshikawa and Sato (2008) found that dynamic presentation of facial emotion creates a perceptual enhancement in the remembered last picture. That is, the picture in front of the inner eye was more intense than the expression presented on the last frame and this effect increased with increasing speed of the developing facial expression of emotion. Within a perceptual enhanced picture of facial emotional expression the features and configurations become more apparent, which might facilitate decoding of facial emotional expressions. Combining the result reported by Yoshikawa and Sato (2008) with the finding that accuracy of response rates increases with increasing expression intensity, this might explain why even subtle displays of facial emotion are recognised significantly above chance level of responding. Also using brain imaging, Yoshikawa and Sato (2006) found dynamic stimuli to elicit perceptual enhancement compared to static facial expressions, which was evident in greater brain activity in brain areas associated with visual perception and emotion processing and therefore relevant to decoding of facial emotional expressions. Precisely, greater brain activation was found in the visual cortex, the amygdala, and the right inferior frontal gyrus in response to dynamic facial expressions than static images of facial emotion. They concluded that stimulus dynamics help to facilitate processing of facial expressions of emotion (see also Sato, Fujimura, & Suzuki, 2008; Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004; Trautmann, Fehr, & Herrmann, 2009). Together, the results reported here suggest that dynamic stimuli of facial emotion intensify the perception and processing of emotions from faces, which might aid the decoding of the observed facial emotional expression. Given these effects of dynamic stimuli and the dynamic nature of facial expressions, it seems imperative to study facial emotion recognition based on video recordings instead of static images or morphed sequences, in order to learn more about facial emotion recognition in the typical population and also the clinical population.

When investigating facial emotion recognition based on static stimuli and of high expression intensity, then existing facial emotion recognition deficits might get missed. For example, deficits in facial emotion recognition that are driven by deficits in motion perception in faces cannot be revealed. Face motion perception deficits presumably become more apparent the smaller the motion is in the face as in subtle facial emotional expressions. This assumption can explain why some behavioural studies did not find facial emotion recognition deficits in ASD given high intensity facial expressions were used especially when based on static images. Since facial expressions are dynamic and often subtle, the finding from the AFDES-BIV (Chapter 8) is in line with reports from individuals with ASD that they have difficulties extracting and interpreting emotional information from faces. The importance of applying video stimuli for assessment of facial emotion recognition ability is thereby demonstrated.

Facial emotion recognition ability across populations.

The studies conducted and presented in this dissertation showed females are better at recognising emotions from faces than males (Chapter 5) and individuals with high-functioning ASD are outperformed by matched typical individuals (Chapter 7). Individuals with ASD are not only outperformed by typical individuals, their recognition ability shows actual deficits contributing to diminished social functioning seen in ASD in everyday life. Facial emotion recognition could be understood as a dimension, with typical females generally having better facial emotion recognition ability than typical males, and typical males having better facial emotion recognition ability than individuals with ASD. This understanding raises the question of whether the facial emotion recognition ability is unique to each population or common to populations. Baron-Cohen (2002) theorises that ASD is the result of an extreme male brain. If that was the case, then the pattern of facial emotion recognition in ASD should be similar to males of the general population, although amplified. Indeed, similar to males performing worse than females at facial emotion recognition across intensity levels, individuals with ASD performed worse than controls at all three intensity levels. Whereas the extreme male brain theory has a biological basis, facial emotion recognition ability can also be explained by a theory based on environmental factors. The social motivation hypothesis was proposed to explain diminished facial emotion recognition abilities in ASD (Dawson et al., 2002; Dawson et al., 2005; Grelotti et al., 2002), but can actually be applied to both populations, male typical individuals and individuals with ASD. Gender-typical socialisation processes might encourage females more than males to engage in social interactions and express feelings and would lead to more experience with emotion recognition in females than males. This greater experience could explain the female superiority in facial emotion recognition. Whereas for males it would be the gender-typical socialisation leading to a disadvantage in facial emotion recognition compared to females, in ASD it would be a preference of self-absorbed activities leading to their poorer performance at facial emotion recognition experiments compared to controls. Since the frequency of encountering specific emotions in daily life has been empirically linked to the recognition of such to explain variations in recognisability between emotions (Calvo et al., 2014), this explanation might apply to the level of inter-individual ability in facial emotion recognition as well. That is, people with more experience in facial emotion recognition are better at facial emotion recognition. This assumption supports the social motivation theory as an explanation for facial emotion recognition ability across populations and implies that facial emotion recognition is trainable, which has implications for interventions for individuals with ASD like social trainings. The extreme male brain theory and the social motivation theory seem applicable for the explanation of facial emotion recognition ability overall.

However, the results from the studies presented within this dissertation do not point towards a commonality of facial emotion recognition ability between populations based on the recognition of individual emotion categories. The facial emotion recognition profile for males was dissimilar to the results from the ASD sample. Taking into account the intensity of emotional facial expression, males also recognised positive emotions worse than females (happiness at intermediate intensity), whereas the ASD group only showed deficits for negative emotions (e.g. fear at all three intensity levels of expression). It seems like poorer facial emotion recognition performance can be seen across populations appearing as a single phenomenon, but the specific profiles of facial emotion recognition actually vary across populations and are each unique to its population. This observation challenges the extreme male brain theory (Baron-Cohen, 2002). Furthermore, this observation is in line with the conclusion drawn by Nuske et al. (2013) that the specific pattern of impairment in ASD is unique to that population based on a literature review of facial emotion recognition ability in ASD compared to other clinical populations, such as individuals with language disorders (Gross, 2004), schizophrenia (Bölte & Poustka, 2003), and ADHD (Demurie, De Corel, & Roeyers, 2011; Sinzig, Morsch, & Lehmkuhl, 2008). With different patterns of facial emotion recognition ability in the varying populations it is likely that distinct processes are underlying the resulting patterns of facial emotion recognition. The literature suggests that for the decoding of facial emotional expressions information is used from several sources, such as facial features, configurations, motion, and velocity. It is likely that deficits or abnormalities in extracting information from any of these sources can result in an emotion recognition deficit with a distinct pattern for each population (which is discussed in more detail below). Future research should seek to compare varying clinical populations next to healthy controls to gain more knowledge about commonalities and specifics in facial emotion recognition ability, since such investigations are sparse.

If different processes are involved in facial emotion recognition, then an important question is whether specific cognitive processes are underlying facial emotion recognition ability (domain-specific) or facial emotion recognition ability is the result of domain-general processes (e.g. attention, working memory, problem solving). For example, it seems as the holistic processing humans apply for face processing is not applied to object processing and the good face recognition does not apply to other body parts (McKone, Kanwisher, & Duchaine, 2007). This specificity for faces indicates domain-specificity for face processing. If facial emotion recognition is domain-specific, then deficits in facial emotion recognition would not influence other abilities. If facial emotion-recognition is dependent on domain-general processes, then deficits in general processes should accompany facial emotion recognition deficits. For example, a general deficit in motion perception would also affect perception of motion in faces (facial expressions). The investigation of confusions in the study presented in Chapter 8 showed that individuals with ASD have deficits in perceiving emotional content from faces, since they perceived disgust, sadness, and anger as neutral to a significantly greater extent than the controls when expressed at low intensity. The literature on motion perception suggests that individuals with ASD have more difficulties in perceiving motion when stimuli are presented briefly (literature review by Dakin & Frith, 2005). As has been discussed in Chapter 7, the exposure time to emotional content was shortest for the low intensity expressions of the ADFES-BIV and only a slight movement was visible. Therefore, it is possible that general deficits in motion perception are what was driving the confusions of emotional facial expressions at low intensity with neutral. This would mean that individuals with ASD have general difficulties in perceiving motion which are not specific to emotional content from faces and that these general motion perception difficulties are causing a facial emotion recognition deficit.

Assuming that domain-general processes underlie facial emotion recognition, it could be presumed that deficits in facial emotion recognition should be seen across emotions instead for specific emotions. However, the results from the studies investigating sex differences and comparing ASD to controls (presented in Chapters 6 and 8) showed reduced recognition for some but not all emotions. The specific emotion deficits can be explained by difficulties in domain-general processes. For example, that only disgust, sadness, and anger at low intensity were perceived as neutral by individuals with ASD and not the other emotions included in the task (results from Chapter 8) can be explained by the results of the study on the intensity judgement of the facial expressions of the ADFES-BIV presented in Chapter 5. The three identified emotions (disgust, sadness, and anger) are the basic emotions that were rated the lowest in intensity by participants. Judgements of low intensity can be interpreted as facial expressions that are more subtle and thereby harder to detect. This difficulty, based on subtlety combined with a motion perception deficit in ASD, could explain why individuals with ASD had difficulties to identify emotional content from anger, disgust, and sadness faces when presented at low intensity. This example explains how domain-general processes can affect facial emotion recognition. The finding used in the example has also strong implications for intervention programs for individuals with ASD, as facial emotion recognition deficits based on perceptual deficits, require different intervention strategies than a facial emotion recognition deficit based on cognitive processes. Future research should investigate motion perception based on dynamic emotional facial expressions of low intensity to clarify the basis of the diminished sensitivity to detect subtle emotional cues in faces. For example, participants (ASD vs controls) could be presented with such stimuli and asked to state whether or not they perceived any movements and if so specify which movements exactly they think to have perceived.

The facial emotion recognition deficit in ASD can further be explained by deficits in general processes applying the weak central coherence theory (Frith, 1989/2003). Accordingly, individuals with ASD focus more on details than on the whole ‘picture’, i.e. they rely more on featural processing than configural processing. For example, if a deficit is present in general configural processing, then the configural processing of faces during decoding of emotional expressions would also be affected. Since some emotions can be recognised with featural processing only (e.g. happiness), the emotion-deficit would only emerge for specific emotions. The deficit in fear recognition that was found within the study in Chapter 8 as well as the high confusion rates of fear with surprise, point towards featural face processing, since for fear recognition configural and holistic face processing is needed whereas surprise can be accurately recognised with featural processing. Sadness and anger were also less well recognised by the ASD group than the controls, which are also emotions that are successfully recognised by employing configural processing of the face. That individuals with ASD have deficits in configural processing not only of faces but also of objects has been demonstrated experimentally. The ASD group was slower in face discrimination and object discrimination compared to controls and showed a preference for local information in two tasks instead of deriving the global whole (Behrmann et al., 2006). Since the detail-focus in ASD seems to apply not only to faces but also to objects, the deficit in facial emotion recognition seems to be the result of domain-general processes.

With individuals with ASD having deficits in motion perception and configural processing, seems as to facial emotion recognition is dependent of domain-general processes. At the same time, individuals with ASD seem to process especially faces differently than controls with reduced fixations to the eyes (Kliemann, Dziobek, Hatri, Steimke, & Heekeren, 2010) and a preference for the mouth region (Spezio, Adolphs, Hurley, & Piven, 2007), which points towards domain-specificity. A preference for the mouth region further suggests that there should not be a facial emotional recognition deficit in ASD for expressions with saliency in the mouth region. Indeed, the results from the study in Chapter 8 did not show deficits in recognition of emotions characterised by salient mouth movements (happiness, pride, contempt, surprise, or disgust) in the ASD group. This suggests that the facial emotion recognition deficit in ASD is largely based on domain-general deficits. Sex differences in facial emotion recognition are more difficult to explain by domain-general processes. For example, differences in motion perception are rarely reported between male and female typical individuals (Handa & McGivern, 2015). In a study that did report a perceptual speed advantage for females compared to males, the found facial emotion recognition advantage for females was still present after controlling statistically for the female perceptual advantage (Hampson et al., 2006). The authors concluded that the female advantage in facial emotion recognition might be an evolved adaptation needed for child rearing. (Whether the female advantage is biologically hard-wired or acquired via socialisation by encouraging females to express and be sensitive to feelings is outside the scope of this dissertation and still a question to be answered). The discussion here suggests that facial emotion recognition underlies domain-specific processes that are not completely independent from domain-general processes. It is possible that facial emotion recognition is domain-specific as suggested by the female advantage that seems to be specific to facial emotion recognition, but that domain-general processes can affect facial emotion recognition. Assuming that individuals with ASD have a facial emotion recognition deficit resulting from domain-specific deficits, which is additionally amplified by deficits in general domain processes, then this could explain why the difference in facial emotion recognition ability between individuals with ASD and controls is as profound as it is.

The effects of proprioceptive feedback on facial emotion recognition.

Patterns of face muscle activation were found with the study presented in Chapter 9 for each of the nine emotions included in the ADFES-BIV that were distinct from the other emotions. Whereas facial mimicry of facial features was evident, the intensity of facial mimicry did not clearly increase form intensity level to intensity level of the ADFES-BIV. In contrast, during explicit imitation the increase in EMG activity from intensity level to intensity level of observed facial expression was extremely prominent. The results suggest that for processing of observed facial emotional expressions the ‘what’ (facial features) is more relevant than the ‘how much’ (intensity). That observed facial features are mimicked but not the intensity of facial feature activation is surprising when assuming that facial mimicry is the result of mirror neuron activity. A literal interpretation of mirror neuron activity would suggest that next to facial features also the relative intensity of the activated facial features is mirrored. Maybe facial mimicry is not based on mirror neuron system activity, but is the result of processing observed facial emotional expressions. It is possible that the processing of observed facial emotional expressions in the visual cortex of the brain extends to motor areas of the brain after the ‘object’ has been identified as a facial expression. In this case, facial mimicry would not constitute a simulation of the observed, but would be the result of emotion processing from faces. More research is needed investigating intensity of facial mimicry in response to varied intensities of observed expression. The study from Chapter 6 found the difference in perceived intensity to be approximately 10% between the intensity levels of the ADFES-BIV. Future research should therefore re-investigate whether intensity of expression is mimicked by employing stimuli of facial emotion at varying intensity levels that are perceived as more different from each other, e.g. 30% intensity for the low intensity expressions, 60% for the intermediate, and 90% for the high intensity expressions.

The suggestion based on the facial mimicry results that the facial features are more important than the intensity of the facial features for processing of observed facial emotional expressions does not fully align with the accuracy rates that were found with the ADFES-BIV across the studies presented in this dissertation. The recognition of individual emotions is dependent on the saliency of facial features and configurations, but intensity of expression affected all emotions resulting in lower recognition rates to low intensity expressions than high intensity expressions. However, the significant interaction of the factors emotion and intensity showed that the importance of intensity of expression varied between emotion categories. Intensity of facial emotional expression was identified as an important factor for recognition (next to features and configurations). These findings suggest that facial mimicry and facial emotion recognition are independent from each other. This assumption challenges the reverse simulation model where facial mimicry is assumed to be a means to facial emotion recognition.

By experimentally manipulating facial movement to investigate the effects on facial emotions recognition, the results from Chapter 9 led to the conclusion that facial mimicry is not a necessity for facial emotion recognition. Facial mimicry had no facilitating effect compared to the blocked mimicry condition for any of the emotions tested. If facial mimicry was necessary for facial emotion recognition, blocking facial mimicry should have led to decreased accuracy rates. That is, facial mimicry does not facilitate facial emotion recognition and can therefore not be a necessity for facial emotion recognition. It seems as to facial mimicry is a by-product of emotion processing rather than a necessity in the recognition process, as also suggested by Rives Bogart and Matsumoto (2010) based on investigation of individuals with face paralysis and their unimpaired emotion decoding ability from faces. However, these results do not necessarily refute the simulation theory as a whole, since facial mimicry did occur, which can be interpreted as evidence for simulation. It is possible that facial mimicry is not part of the decoding process even though part of the simulation. Though, that the intensity of observed facial expressions did not reflect relatively in the facial mimicry challenges this possibility.

In contrast to the lack of influence of facial mimicry on facial emotion recognition, explicit imitation had some facilitating effects. Disgust recognition was improved when participants imitated the expression compared to when facial movements were blocked. The recognition of embarrassment was improved during explicit imitation compared to the no face movement manipulation condition. That there is an effect of proprioceptive feedback and disgust recognition is in line with published reports. Regarding the influence of proprioceptive feedback and disgust recognition, Oberman et al. (2007) reported similar findings in that they found recognition of disgust to be diminished in the blocking condition compared to the no face movement manipulation condition. They interpreted this finding as facial mimicry being necessary for facial emotion recognition. However, this finding can also be interpreted the other way around, in that incompatible muscle activations hamper recognition. In the current study, disgust can be expected to be affected greatest by the blocking of all emotions, since muscles were activated by the pen holding that are antagonists to the levator. That is, muscular movements of the levator were fully blocked, as putting pressure on a pen with the lips is incompatible with wrinkling the nose. It can be concluded that when muscular feedback is fully blocked by activation of antagonist muscles, recognition is diminished. Regarding embarrassment, it is possible that explicit imitation of the facial expression of embarrassment and the resulting enhanced proprioceptive muscle feedback made the expression less ambiguous, which helped recognition. In contrast to during disgust recognition, the pen holding did not diminish the accuracy rates for embarrassment which might be due to the fact that the depressor was highly activated during blocking and therefore facilitated instead of hampered embarrassment recognition. This is based on the finding that embarrassment was identified as the emotion with most depressor activity during spontaneous facial mimicry compared to the other muscles. Overall, the results from the study presented in Chapter 9 show that with the observed expression incompatible muscle activations hamper recognition and compatible muscle activations facilitate recognition, although this was only found for disgust as that was the only emotion for which facial features were fully blocked. This finding aligns with the reports by Schneider et al. (2013), as they found participants to achieve higher recognition rates when explicitly imitating compared to suppressing imitation. The findings also align with the wider literature from embodied cognition where bodily actions/muscle activations influence judgements of a stimulus (e.g. Strack et al., 1988). It seems like enhanced proprioceptive feedback can facilitate recognition, but being able to engage in facial mimicry does not have facilitating effects.

The fact that facial mimicry did occur but had no effect on facial emotion recognition suggests that visual information about the features and configurations seem to be more important during decoding of facial emotion than proprioceptive feedback from facial mimicry. That is not to say that face muscle activations do not have any influence. Face muscle activations can influence recognition, but only for larger muscle activations like when explicitly imitating, although the found effects were small. Subtle muscle activations as measured via EMG during spontaneous facial mimicry seem to not have an effect on recognition rates. It is also possible that the muscle activity during facial mimicry could have an effect if it was conscious like explicit imitation is. It would be interesting to investigate recognition after participants have been informed about the occurrence of facial mimicry. Future research could also test for effects of incompatible facial mimicry by having participants look at facial expressions leading to facial mimicry and present a different facial emotional expression in the peripheral visual field, which needs to be recognised. The results should then be contrasted to an experimental condition with congruent emotion displays.

Future research should further aim to develop a method to block facial mimicry in the whole face to investigate the effect on facial emotion recognition. Holding a pen in the mouth leaves the muscles of the upper face unaffected. This partial blocking of facial mimicry can therefore only have a limited effect on the accuracy rates. It should further be systematically tested what effect enhanced proprioceptive feedback has on facial emotion recognition. This could be achieved by instructing the participant to activate a specific muscle and investigate the effect on recognition of emotions that involve antagonist muscles. Participants could be asked to smile, frown, wrinkle the nose etc. across a set amount of trials displaying varying emotions and investigated if incompatible facial movements decrease recognition compared to compatible expressions.

There has been a learning outcome from the study presented in Chapter 9 on a methodological level, which should be taken into account when assessing face EMG. During data assessment of facial mimicry, it would be beneficial to ask participants to perform a maximal activation of the face regions measured. This way, EMG responses could get related to the maximum value. Every response to a stimulus would then constitute a proportion out of the maximum and properly reflect the intensity of facial mimicry.

Strengths and weaknesses of the ADFES-BIV and directions for future work.

The strengths and weaknesses of the studies conducted within the current research project were discussed in the individual study chapters and are thus not repeated here. Limitations and delimitations surrounding the participants, the recruitment of participants, lab-based assessment of facial emotion recognition etc. were discussed in Chapter 3. These are thus also not repeated here. Instead, since the stimuli were applied across all investigations conducted within the current research project were from the same face emotion stimulus set, the strengths and weaknesses of the ADFES-BIV are discussed in the following paragraphs. In addition, suggestions for future research are given.

The development of face emotion stimulus sets started off with static images of high intensity facial expression and then moved to variations of expression intensity created by applying the morphing technique. These computer-generated images were then displayed after each other resulting in dynamic sequences. Static images of high expression intensity displaying posed emotions have low ecological validity, and live interactions where emotions are displayed as they are felt (not posed) including varying expression intensities, have the highest ecological validity. Emotion researchers currently advocate the use of dynamic stimuli instead of static images. Some concluded that stimuli of dynamic facial emotional expressions might be more appropriate to use than static stimuli when investigating emotion processing from faces (Trautmann et al., 2009). There is a plethora of research based on static images of high facial emotional expression intensity, which has advanced our understanding of emotions. However, our knowledge and understanding of the processing of subtler emotional displays is limited due to the low number of studies applying intensity variations of facial emotional expressions. With the development of the ADFES-BIV it was desired to include variations in expression intensity and motion, but without computer-generating these variations to include some characteristics contributing to ecological validity.

It is a strength of the ADFES-BIV is that it comprises of video stimuli. Video stimuli provide cues that facilitate facial emotion recognition that are not included in static images, since video stimuli include motion providing a perception of change. A perception of change can also be provided by presenting two static images, for example how it was done in the study in Chapter 7 (first-last frame stimuli). Although the first-last frame stimuli provided a perception of change – the change from neutral to the emotional expression – the emotion-specific temporal characteristics containing important information for decoding (Ekman & Friesen, 1982; Hess & Kleck, 1990) were stripped. For example, slowness could be the temporal characteristic of sadness. Video recordings of an unfolding facial emotional expression contain these temporal characteristics. The temporal information is used during the decoding process of the observed facial emotion. Static images do not contain temporal characteristics about emotion by nature and even in morphed sequences this is not guaranteed as the researcher determines the frame rate and thereby how fast the facial emotional expression unfolds. That the accuracy rates from the study in Chapter 7 based on the first-last frame videos were slightly lower than in the other studies demonstrates the importance of motion/dynamics for recognition (Bold et al., 2008, Bould & Morris, 2008). During the development of the ADFES-BIV the frame rate was unaltered allowing for the advantage of offering the temporal characteristics of emotional expressions as recorded. It can be concluded video stimuli the perception of motion and the motion itself contain emotional information demonstrating the necessity to employ video stimuli for investigations on processing of facial emotion. Krumhuber et al. (2013) concluded in a recent literature review that next to limited ecological validity static expressions restrict our understanding of facial emotional expressions. It can be concluded that research on facial emotion recognition should utilise dynamic face stimuli.

Another strength of the ADFES-BIV is the inclusion of varying levels of expression intensity. Although variations in expression intensity have been used based on computer-morphs, the discussion about emotion-specific temporal characteristics demonstrates the importance to investigate the processing of subtler facial emotional displays based on video recordings. The inclusion of varying levels of expression intensity allows for fine-graded analysis of emotion processing from faces. Intensity of expression influenced the recognisability of the emotional facial expressions across the studies conducted within the current research project. This influence did not only vary between emotions, but also between participant groups. The findings from the studies presented within this dissertation showed that knowledge can be gained advancing the literature. For example, it became apparent that deficiencies in multiple processes are underlying the facial emotion recognition deficits in ASD by investigating recognition and error rates across intensity levels. The unexpected finding from the study presented in Chapter 9 that the intensity of observed facial emotional expression did not reflect in the intensity of facial mimicry showed that more attention should be paid to facial responses to subtle facial expressions. It would be interesting to investigate brain processing of subtler displays of facial emotion. A potential research question is whether the intensity of observed facial emotional expressions modulates the resulting amount of brain activation, which the ADFES-BIV offers the possibility to investigate.

A limitation of the ADFES-BIV is that the intensity levels were achieved via an editing process where video recordings were truncated. With this approach the offsets of the emotional displays were not visible. That is, the videos started with a neutral facial expression that then developed into a facial expression of emotion (therefore including the onset of the emotional facial expression), but the end point of each video is an emotional expression at either low, intermediate, or high expression intensity. In social interactions, the offset of the facial emotional expression is generally visible as well. Emotional facial expressions generally reach their apex and remain there for a short amount of time before the expression develops back to a neutral face or into another emotion. Future research should record emotions at varying intensities showing onsets and offsets of the emotional facial expressions. This would make it necessary to instruct individuals to portray emotions at varying intensities. It is possible though that portraying a subtle emotion might be difficult to execute for naïve individuals on command, as the obvious question would be what constitutes subtle. Professional actors could be used to portray emotions in that matter. The downside of this approach is that actors are trained to portray emotions in a specific, often exaggerated manner, to make it understandable for the audience. Portrayals of emotion in by an untrained person might deviate from portrayals by actors.

There seems to be a trade-off between standardisation and ecological validity when developing a stimulus set of facial emotion. For example, the inclusion of varying levels of expression intensity in the ADFES-BIV led to variations in the amount of time emotional information is visible in the videos. This fact made it necessary to conduct a study with controlled timings to investigate the potential effects of variations in the amount of time emotional information is visible in the videos on facial emotion recognition. The more ecological valid stimuli, the more difficult it is to reach standardisation. Most face emotion stimulus sets include posed facial expressions, although elicited emotional expressions would be higher in ecological validity. The difficulties with eliciting emotions are: to first of all elicit an emotion, to elicit the intended emotion, to witness a facial expression that is neither masked or suppressed in any way, and to standardise the expressions in terms of angle, activated facial features, head movements, timings etc. Nonetheless, future research should attempt to develop a stimulus set including elicited emotions of facial expressions with the highest possible degree of standardisation. Our knowledge and understanding of facial emotional expressions can get advanced by comparing results from stimulus sets of higher ecological validity to the results from more standardised stimulus sets.

This discussion demonstrates that there are further obstacles that need to get tackled when continuing the production of more ecological valid stimuli for application within emotion processing research. The outlined difficulties that come with any possible approach when creating such stimuli further demonstrate the value of the ADFES-BIV, despite its limitations.

Conclusion

A video stimulus set of facial emotion including varying levels of expression intensity was developed and validated. Accuracy rates well above chance level of responding were retrieved with the ADFES-BIV and it proved suitable for investigation of group differences in facial emotion recognition. The ADFES-BIV was found as versatile and suitable for research on emotion processing from emotion recognition to emotion production in typical and clinical samples like ASD within the current research project. By applying the ADFES-BIV in the general population, a female advantage over males in facial emotion recognition was found. Within the clinical population, individuals with ASD were found to have deficits in facial emotion recognition that are consistently evident across intensity levels. Across all studies conducted, it was shown that the difficulty to recognise a facial expression of emotion is determined by the emotion category itself, but also by the intensity of the facial expression. It was further shown that facial emotion recognition is linearly related to difficulty based on accuracy of response. This finding could not be explained by the facial mimicry response. Whereas facial features are being mimicked, the intensity of the features seems to not be part of facial mimicry. The linear relationship between facial emotion recognition and difficulty also translates to the pattern of group differences in facial emotion recognition ability across expression intensity levels. The investigation of facial mimicry showed distinct patterns of facial muscle activation for all emotion categories of the ADFES-BIV. Overall, facial mimicry seems to be more of a by-product of emotion processing than a tool to aid facial emotion recognition, although intensified proprioceptive feedback does have some effects in line with embodied cognition. Whereas incongruent muscle activation can hamper recognition, facilitating effects can occur during congruent muscle activation. Facial emotion recognition seems to be the result of a multitude of processes, some of which are specific to faces (e.g. holistic processing) and some are general processes (e.g. motion perception).

APPENDICES

Appendix A. Pre-study: neutral clip selection

Aims and Hypothesis

The aim of this study was to identify a video clip of neutral content. Based on ratings before and after each of the clips included in this study, the clip with the ratings closest to neutral after watching would be selected. Considered as neutral were a valence rating around 3 (= neutral) and arousal around 2 (= low) in the post-rating.

Method

Participants.

Participants were recruited through the University of Bath noticeboard and the Department of Psychology Research Participation Scheme. The sample consisted of 12 participants (10 females, 2 males), aged between 18 and 27 (age; M = 20.4, SD = 3.0); nine 1st year Psychology undergraduate students who gained course credits for their participation and 3 Psychology PhD students who participated in exchange for the author's participation in their research. The sample was culturally diverse: British (n = 7), American, Chinese, Cypriot, Vietnamese, and British/Irish (n = 1 each) with all participants assumed to be fluent in English. None of the participants reported a current diagnosis of a mental disorder. Ethical approval for this study was obtained from the University of Bath, Department of Psychology Ethics Committee prior to the start of the investigation (REF 13-044).

Stimuli.

Based on a subjective judgement, four different video clips were pre-selected by the author to test for being neutral in content. Next to the content, the clips were selected on the basis of their duration, which was required to be between (approximately) five and ten minutes. This video clips were short documentaries about various topics: 1) "The naked eye" (6:59min); an astronomy documentary describing stars which are visible to us without telescope (from now on referred to as Astro). 2) “The ghost-town project” (10:08min); a BBC documentary about a graffiti art project in an abandoned town (from now on referred to as Ghost). 3) “Giant rodents” (3:54min); a David Attenborough BBC documentary about giant rodents (from now on referred to as Rodents). 4) “Fertilizers” (4:18min); a documentary explaining the soil-plant-nutrient-relationship.

Ratings.

The participants’ affective state were assessed through the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), a non-verbal measure on a 5-point scale of arousal (1 = very weak, 2 = weak, 3 = moderately, 4 = strong, 5 = very strong) and valence (1 = very positive, 2 = positive, 3 = neutral, 4 = negative, 5 = very negative).

Procedure.

The study was conducted in a Psychology laboratory of the University of Bath. Participants were tested one after the other in presence of the experimenter. Participants were informed that their participation is voluntary and they could stop the experiment at any time without providing a reason. All participants gave written informed consent prior to participation and were fully debriefed on completion of the study. The experiment was implemented in Microsoft PowerPoint (2010). A within-subject design was applied. The presentation sequence was alternated to avoid sequence effects and participants randomly assigned. The experiment started with the ratings on valence and arousal before presentation of the videos and ratings after each presentation on the same two items. Accordingly, ratings took place five times. Duration of the experiment was approximately 30 minutes.

Data preparation and analysis.

Analyses were conducted using SPSS version 21 (SPSS IBM, New York, USA). Due to its suitability for paired sample comparisons of ordinal scaled variables Wilcoxon matched pairs signed-rank tests were applied to compare the pre- and-post-ratings of valence and arousal. Although non-parametric tests are robust against outliers, the data was inspected by boxplots to assure there were no systematic outliers. Only three univariate outliers were identified (for ‘Ghost’ on valence, subject 9, 6, & 12) due to the other participants all rating the same value. The identified data points remained included in the analyses.

To test whether the Wilcoxon matched pairs signed-rank test’s requirement of symmetrical distribution of the differences of the scores on each variable was met, compound variables were created (value pre-rating - post-rating). According to the suggestion of Field (2009) for small sample sizes skewness and kurtosis -values were examined to check for normality in the distribution of the data where a non-normal distribution is defined as standardised skewness and kurtosis values greater 1.96.

The valence ratings ‘Astro - pre-rating, ‘Rodents - pre-rating’, and ‘Fertilisers - pre-rating’ seemed normally distributed with standardised skewness and kurtosis values < 1.96. The valence ratings of ‘Astro - pre-rating’ were negatively skewed (-1.60) with a standardised skewness of 2.51 as well as leptokurtic (3.83) with a standardised kurtosis of 3.11. All arousal ratings seemed normally distributed with standardised skewness and kurtosis values < 1.96. Based on the results a symmetrical distribution of the differences of the scores on each (but one) variable was met and therefore Wilcoxon matched pairs signed-rank tests applied. Exact *p*-values were obtained due to the small sample size, as recommended by Field (2009). None of the participants suspected the aim of the study, to identify the video which is the most neutral. However, one participant said that all video sequences appeared quite neutral to them. Analyses were conducted including and excluding this participant and as results were not influenced, the participant remained included for analysis.

Results

Valence.

Wilcoxon signed-rank tests showed the median valence scores from the initial rating (*Mdn* = 4.00) decreased significantly after watching ‘Ghost’ (*Mdn* = 3.00, z = -2.75, exact p = .004, r = -.56) and ‘Fertilisers’ (*Mdn* = 3.00, z = -2.48, exact p = .018, r = -.51). After watching ‘Rodents’ (*Mdn* = 4.00, z = -.71, exact p = .750) and ‘Astro’ (*Mdn* = 4.00, z = -1.41, exact p = .312) no changes occurred in the valence ratings.

The mean valence rating from before the experiment was 4.08 (*SD* = 0.67). After watching ‘Fertilisers’ the mean valence rating decreased to 3.17 (*SD* = 0.58). After watching

‘Astro’ the mean valence rating was 3.67 (*SD* = 0.65), after ‘Rodents’ the mean valence rating was 4.25 (*SD* = 0.62), and after ‘Ghost’ the mean valence rating was 2.92 (*SD* = 0.52). ‘Ghost’ and ‘Fertilisers’ approach closest the intended score of 3 on valence. Additionally, one sample *t*-tests were conducted to test whether the valence scores after the clips were significantly different from the desired score of 3. Results showed that ‘Fertilisers’ was not significantly different from the value 3 ($t(11) = 1.00$, $p = .339$), as so ‘Ghost’ ($t(11) = -0.56$, $p = .586$). ‘Rodents’ was flagged significantly different from the value 3 ($t(11) = 6.97$, $p < .001$), as well as ‘Astro’ ($t(11) = 3.55$, $p = .005$).

Arousal.

The arousal rating decreased significantly from the initial rating (*Mdn* = 3.00) compared to the rating after ‘Fertilisers’ was watched (*Mdn* = 1.50, $z = -2.97$, exact $p = .002$, $r = -.60$). The changes in arousal after ‘Astro’ (*Mdn* = 2.00, $z = -2.12$, exact $p = .055$), ‘Ghost’ (*Mdn* = 2.00, $z = -1.22$, exact $p = .305$), and ‘Rodents’ (*Mdn* = 3.00, $z = -.71$, exact $p = .750$) were not found to be significant by Wilcoxon signed-rank tests.

The mean arousal rating from before the experiment was 2.75 (*SD* = 0.62). After watching ‘Fertilisers’ the mean arousal rating was 1.75 (*SD* = 0.87). After watching ‘Astro’ the mean arousal rating was 2.00 (*SD* = 0.74), after ‘Rodents’ the mean arousal rating was 2.58 (*SD* = 0.79), and after ‘Ghost’ the mean arousal rating was 2.25 (*SD* = 1.22). One sample *t*-tests showed arousal ratings were significantly different from the desired value of 2 on arousal for ‘Rodents’ ($t(11) = 2.55$, $p = .027$) and not significantly different for ‘Fertilisers’ ($t(11) = -1.00$, $p = .339$), ‘Ghost’ ($t(11) = 0.71$, $p = .491$), and ‘Astro’ ($t(11) = 0.00$, $p = 1.00$).

Conclusion

Results showed that ‘Fertilisers’ and ‘Ghost’ were of neutral content based on measurement of self-rated affective state on valence and arousal. It can therefore be concluded that the two clips are of neutral valence and arousal. Due to the shorter duration it was decided to apply ‘Fertilisers’ within the studies conducted of the current research project.

Appendix B. GLMM with accuracy of response data uncorrected for extreme values.

This analysis refers to Chapter 6. The main effect of *emotion* was significant ($F(8,86) = 56.05$, $p < .001$). Pairwise contrasts showed that only anger and sadness, contempt and pride, and disgust, embarrassment, and fear did not differ significantly in their accuracy rates from each other (p's $> .05$, after sequential Bonferroni-correction). All other emotions differed significantly from each other based on accuracy or response rates (p's $< .05$, after sequential Bonferroni-correction).

The main effect of *intensity* was significant ($F(2,639) = 403.30$, $p < .001$). Pairwise contrasts showed that all three intensity levels were significantly different from each other based on accuracy of response rates (p's $< .001$, after sequential Bonferroni-correction).

The interaction of *emotion*intensity* was significant ($F(16,578) = 21.06$, $p < .001$). Pairwise contrasts showed that only the recognition rate of sadness at intermediate expression intensity was not significantly different from sadness expressed at high intensity ($p = .164$, after sequential Bonferroni-correction). For surprise, the differences between the recognition rates at low intensity to intermediate intensity and from intermediate intensity to high intensity were flagged as trends (p's $= .066$, after sequential Bonferroni-correction). For all other emotion categories the differences between the three intensity levels were significant (p's $< .05$, after sequential Bonferroni-correction).

The main effect of *sex* was significant ($F(1,95) = 13.11$, $p < .001$). Simple contrasts showed that after sequential Bonferroni-correction females achieved significantly higher accuracy rates than males ($p = .001$).

The interaction of *sex*emotion* was not significant ($F(8,86) = 1.50$, $p = .171$).

The interaction of *sex*intensity* showed a trend ($F(2,639) = 2.73$, $p = .066$). Simple contrasts showed that females achieved significantly higher recognition rates at all three levels of expression intensity but the difference was most prominent at intermediate intensity of expression (p's $\leq .01$, after sequential Bonferroni-correction).

The interaction of *sex*emotion*intensity* showed a trend ($F(16,578) = 1.57$, $p = .072$). Simple contrasts showed that females recognised the emotions anger, disgust, embarrassment, and sadness significantly better than males at low expression intensity (p's $< .05$, after sequential Bonferroni-correction). At intermediate expression intensity females recognised the emotions anger, contempt, disgust, embarrassment, fear, happiness, and sadness significantly better than males (*p*'s < .05, after sequential Bonferroni-correction). At high intensity females recognised the emotion anger and contempt significantly better than males (*p*'s < .01).

Appendix C. GLMM with response latencies data uncorrected for extreme values.

This analysis refers to Chapter 6. Since only correct trials were included in the analysis, 4.1% of the data at category level was excluded. The main effect of *emotion* was significant ($F(8,149) = 23.50$, $p < .001$). Pairwise contrasts showed that responses occurred significantly faster to happiness than anger, disgust, fear, sadness, embarrassment, and contempt (p's < .05, after sequential Bonferroni-correction). Responding occurred significantly faster to surprise than anger, disgust, fear, embarrassment, and contempt (p's < .05, after sequential Bonferroni-correction). The responses to sadness, disgust, anger, and pride were significantly faster than to fear and contempt (p's < .05, after sequential Bonferroni-correction). Responding occurred significantly faster to embarrassment and fear than contempt ($p < .001$, after sequential Bonferroni-correction); a trend was found for faster responses to embarrassment than fear ($p = .076$, after sequential Bonferroni-correction) and sadness ($p = .051$, after sequential Bonferroni-correction).

The main effect of *intensity* was significant ($F(2,266) = 80.18$, $p < .001$). Pairwise contrasts showed that all three intensity levels were significantly different from each other based on response latencies (p's < .001, after sequential Bonferroni-correction).

The interaction *emotion*intensity* was significant ($F(16,290) = 5.89$, $p < .001$). Pairwise contrasts showed that responding occurred significantly faster to intermediate intensity anger expressions than low anger expressions ($p < .001$, after sequential Bonferroni-correction) and a trend was found for the difference between intermediate and high intensity of anger ($p = .070$, after sequential Bonferroni-correction). For disgust, responses occurred significantly faster to high intensity expressions than intermediate intensity expressions ($p = .005$, after sequential Bonferroni-correction) and a trend was found for the difference between intermediate and low intensity ($p = .081$, after sequential Bonferroni-correction). For fear, responses occurred significantly faster to high intensity expressions than intermediate intensity expressions ($p < .001$, after sequential Bonferroni-correction), the difference between intermediate and low intensity was not significant ($p = .115$, after sequential Bonferroni-correction). For sadness, responses occurred significantly faster to intermediate expressions than low intensity expressions ($p = .034$, after sequential Bonferroni-correction), the difference between intermediate and high intensity expression was not significant ($p = .568$, after sequential Bonferroni-correction). The differences in response latencies between the intensity levels of surprise were all significant (p's < .01, after sequential Bonferroni-correction), as well as for happiness (p's < .001, after sequential Bonferroni-correction) and embarrassment (p's < .05, after sequential Bonferroni-correction). The differences in response latencies between low and intermediate pride expression showed a trend ($p = .059$, after sequential Bonferroni-correction) and no significance emerged for intermediate and high intensity of pride ($p = .803$, after sequential Bonferroni-correction). The intensity levels of contempt were not significantly different from each other (p's > .05, after sequential Bonferroni-correction).

The main effect of *sex* was significant ($F(1,114) = 6.44$, $p = .013$). Simple contrasts showed that females responded significantly faster than males ($p = .014$, after sequential Bonferroni-correction).

The interaction *sex*emotion* was not significant ($F(8,149) = 0.99$, $p = .447$).

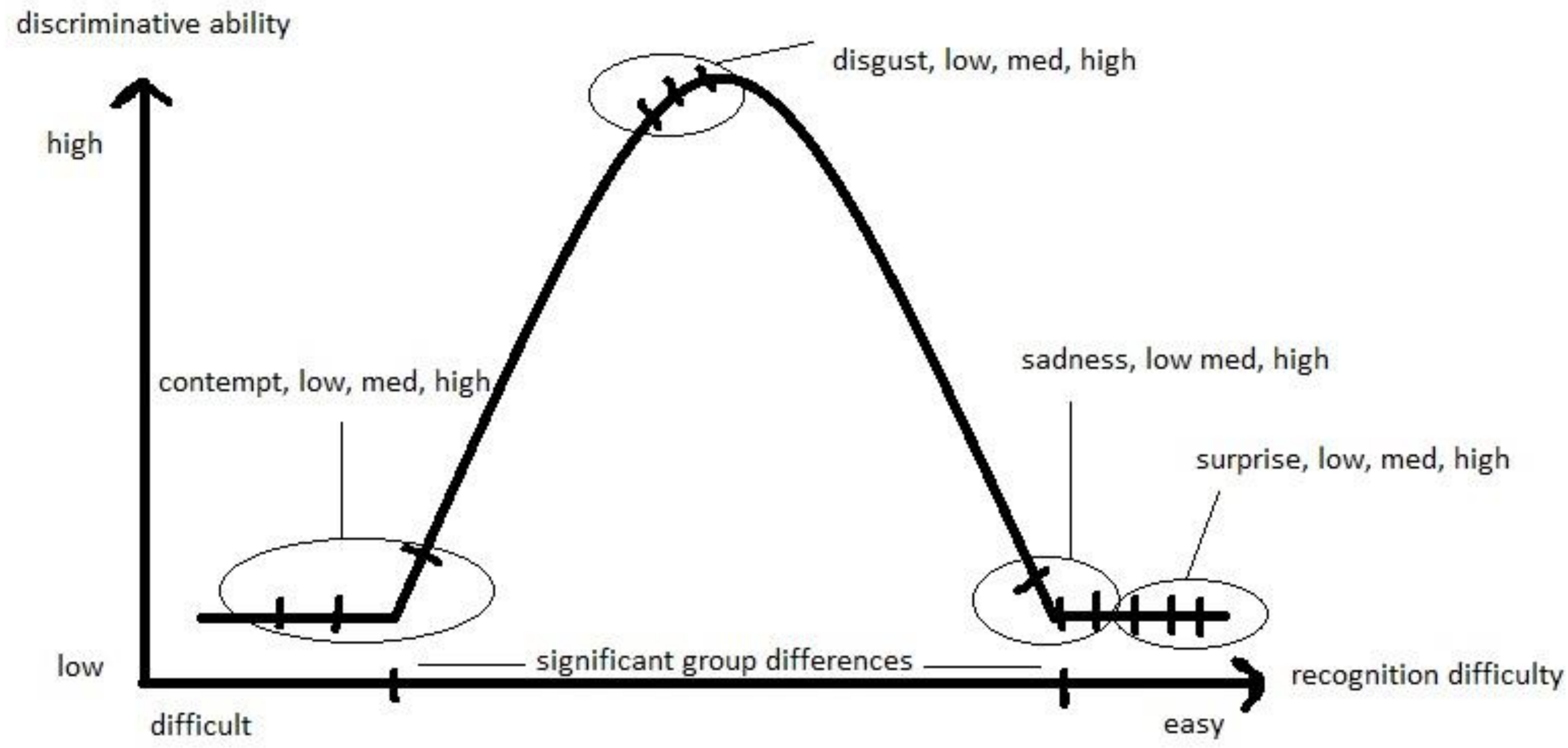
The interaction *sex*intensity* was not significant ($F(2,266) = 1.21$, $p = .300$).

The interaction *sex*emotion*intensity* showed a trend ($F(16,290) = 1.60$, $p = .067$). Simple contrasts showed that females recognised the emotions disgust, happiness, and sadness significantly faster than males at low expression intensity (p's < .05, after sequential Bonferroni-correction). At intermediate expression intensity females recognised the emotions disgust, happiness, and sadness significantly faster than males (p's < .05, after sequential Bonferroni-correction). At high intensity females recognised the emotion happiness and contempt significantly faster than (p's < .05, after sequential Bonferroni-correction). A trend was found for females responding faster than males to anger at intermediate intensity ($p = .074$, after sequential Bonferroni-correction).

Appendix D. Accuracy of response per video.

		Encoder											
Emotion	**Intensity**	**F01**	**F02**	**F03**	**F04**	**F05**	**M02**	**M03**	**M04**	**M06**	**M08**	**M11**	**M12**
anger	low	.52	.65	.32	.72	.76	.35	.75	.59	.42	.74	.71	.62
	med	.54	.86	.86	.93	.80	.60	.79	.70	.90	.77	.88	.80
	high	.70	.89	.92	.93	.95	.80	.97	.51	.96	.83	.93	.82
contempt	low	.23	.37	.33	.30	.13	.17	.24	.29	.27	.14	.33	.45
	med	.30	.46	.25	.37	.25	.23	.43	.36	.39	.40	.47	.51
	high	.42	.53	.36	.41	.30	.33	.42	.33	.46	.39	.41	.55
disgust	low	.72	.66	.17	.84	.64	.58	.59	.45	.55	.62	.67	.52
	med	.70	.59	.42	.89	.68	.76	.61	.50	.77	.62	.66	.58
	high	.92	.57	.46	.84	.75	.87	.68	.58	.77	.71	.68	.72
embarrassment	low	.62	.29	.84	.82	.34	.32	.46	.35	.63	.41	.37	.13
	med	.80	.51	.91	.91	.68	.80	.61	.38	.67	.61	.40	.25
	high	.79	.77	.96	.88	.89	.89	.88	.90	.75	.72	.79	.91
fear	low	.32	.72	.65	.65	.46	.48	.48	.43	.42	.50	.37	.70
	med	.79	.79	.72	.70	.47	.58	.62	.50	.65	.54	.46	.72
	high	.83	.82	.78	.66	.84	.55	.67	.59	.72	.70	.68	.72
happiness	low	.82	.65	.74	.67	.35	.66	.72	.82	.82	.49	.64	.80
	med	.92	.87	.99	.84	.95	.89	.89	.98	.96	.77	.86	.83
	high	.99	.98	.99	.91	.99	1.00	.97	.99	.97	.93	.84	.95
neutral	1	.75	.95	.87	.96	.68	.93	.98	.95	.96	.70	.89	.88
	2	.92	.95	.92	.97	.93	.88	.90	.95	.98	.72	.83	.84
	3	.80	.93	.83	.98	.90	.88	.96	.92	.98	.84	.87	.88
pride	low	.36	.29	.46	.21	.35	.25	.26	.12	.05	.60	.40	.20
	med	.51	.47	.53	.52	.48	.28	.29	.36	.42	.63	.48	.40
	high	.59	.48	.61	.60	.55	.36	.36	.45	.54	.62	.52	.57
sadness	low	.89	.70	.76	.88	.46	.33	.83	.49	.61	.90	.93	.89
	med	.98	.83	.72	.93	.55	.28	.96	.89	.88	.95	.93	.91
	high	.97	.77	.83	.95	.54	.61	.89	.89	.86	.91	.88	.92
surprise	low	.78	.95	.79	.91	.96	.90	.87	.96	.93	.95	.89	.96
	med	.91	.83	.92	.88	.92	.92	.88	.99	.95	.91	.98	.95
	high	.95	.90	.91	.96	.99	.93	.93	.93	.93	.99	.96	.95

Appendix E. Relationship between recognition difficulty and discriminative ability.



Exemplification of the relationship between the difficulty to recognise a facial expression of emotion and the discriminative ability in detecting group differences in facial emotion recognition (here: sex differences).

Appendix F. Pairwise comparisons from Chapter 8.

Pairwise comparisons for the significant main effects of *emotion* and *intensity* and the significant interaction of *emotion*intensity* from the study presented in Chapter 8 (ASD vs controls).

The emotions were recognised in the following order (from highest to lowest accuracy rates): surprise ($M = .94$, $SD = .02$), happiness ($M = .89$, $SD = .02$), sadness ($M = .79$, $SD = .02$), disgust ($M = .73$, $SD = .05$), anger ($M = .71$, $SD = .05$), embarrassment ($M = .60$, $SD = .05$), fear ($M = .51$, $SD = .05$), pride ($M = .48$, $SD = .06$), contempt ($M = .34$, $SD = .07$). Surprise and happiness were significantly better recognised than all other emotions (p's < .05, after sequential Bonferroni-correction), but were not significantly different from each other ($p = .331$). Sadness was significantly better recognised than fear ($p < .001$), contempt ($p < .001$), embarrassment ($p = .003$), and pride ($p < .01$), but not anger ($p = .488$) and disgust ($p = .764$); all p's after sequential Bonferroni-correction. Disgust was significantly better recognised than contempt ($p < .001$), pride ($p < .001$), and fear ($p = .013$), but not anger ($p = 1.00$) and embarrassment ($p = .230$); all p's after sequential Bonferroni-correction. Embarrassment was significantly better recognised than contempt ($p = .002$), but not fear ($p = .515$) and pride ($p = .260$); all p's after sequential Bonferroni-correction. Recognition of fear was not significantly different to contempt ($p = .266$) and pride ($p = 1.00$); all p's after sequential Bonferroni-correction. Recognition of pride was not significantly different to contempt ($p = .088$, after sequential Bonferroni-correction).

Pairwise contrasts showed that the accuracy rates for the high intensity expressions ($M = .80$, $SD = .02$) were significantly higher than for the intermediate intensity expressions ($M = .71$, $SD = .03$; $t(594) = 6.79$, $p < .001$, after sequential Bonferroni-correction), and for the latter significantly higher than for the low intensity expression ($M = .56$, $SD = .03$; $t(594) = 8.69$, $p < .001$, after sequential Bonferroni-correction).

Pairwise contrasts showed that for most emotions the accuracy rates were significantly higher at high intensity than at intermediate intensity and significantly higher at intermediate intensity than at low intensity (p's < .05, after sequential Bonferroni-correction); see Table below. Only for surprise there were no significant differences between the intensities (p's > .05, after sequential Bonferroni-correction). For anger and sadness the difference in accuracy of response was not significant between the intermediate and high intensities (p = .489, after sequential Bonferroni-correction). For disgust the difference in accuracy of response was not significant between the low and intermediate intensities (p = .214, after sequential Bonferroni-correction). For contempt only the difference between the low and high intensity was significant (p = .001, after sequential Bonferroni-correction).

Table. Means and Standard Errors of the Means for the Emotions of the ADFES-BIV at Each Intensity Level (Chapter 8).

Emotion	**low**		**intermediate**		**high**	
(*n* = 12)	***M***	***SE***	***M***	***SE***	***M***	***SE***
Anger	.50	.06	.72	.04	.83	.04
Sadness	.69	.03	.82	.02	.84	.02
Disgust	.66	.05	.70	.06	.81	.04
Fear	.38	.04	.51	.05	.63	.05
Happiness	.70	.05	.90	.02	.96	.01
Surprise	.92	.02	.94	.01	.94	.02
Contempt	.30	.06	.34	.07	.41	.08
Embarrassment	.38	.04	.57	.05	.81	.04
Pride	.34	.06	.50	.07	.61	.06

Note. Means (*M*) and standard errors of the means (*SE*) are expressed in decimal fractions.

Appendix G. Within-task improvements in accuracy of response.

It was explored whether or not there would be any changes in accuracy of response over the course of the experiment based on non-specific *improvement effects*. It was assumed that accuracy rates would increase over the course of the experiment.

Data preparation and analysis.

To investigate the accuracy of response data for learning effects within the experiment, the 360 trials were divided in four blocks with 90 trials each, in consecutive order (as participants underwent them). A division in four blocks was chosen, since the study presented in Chapter 9 consisted of four parts (experimental conditions) with 90 videos each, so that the results of the current study could inform the study in Chapter 7. Whereas the data for blocks 1-3 showed a normal distribution (W's(92) > .98, p's > .05), the data for block 4 was significantly different to a normal distribution ($W(92) = .97$, $p = .047$), left-skewed. Square transformation was therefore conducted for all four blocks to normalise the distribution of the four variables. Shapiro-Wilk tests of the transformed data showed that the data were successfully normalised (W's(92) > .99, p's > .05). No extreme values were identified using boxplots. The accuracy scores between the four blocks were compared with paired sample t-tests. A Bonferroni-corrected p-value of .008 ($p = .05/6$) was applied to account for the multiple comparisons.

Results.

Paired sample t-tests comparing the blocks including 90 videos each to each other showed that there was a significant increase in accuracy of response from block 1 ($M = .64$, $SE = .01$) to block 2 ($M = .68$, $SE = .01$) of 4% ($t(91) = -4.11$, $p < .001$) and a significant increase of 2% from block 2 to block 3 ($M = .71$, $SE = .01$, $t(91) = -3.13$, $p = .002$). The increase of 1% from block 3 to block 4 ($M = .72$, $SE = .01$) was not significant ($t(91) = -1.44$, $p = .155$) meaning that the improvements in accuracy were strongest over the first three quarter of the experiment (270 trials). However, the difference between block 1 and block 4 was highly significant ($t(91) = -7.30$, $p < .001$), as so was the difference between block 1 and block 3 ($t(91) = -6.25$, $p < .001$), 2 and 4 ($t(91) = -4.79$, $p < .001$); see Figure below. The total increase in accuracy from block 1 to block 4 was 8%.

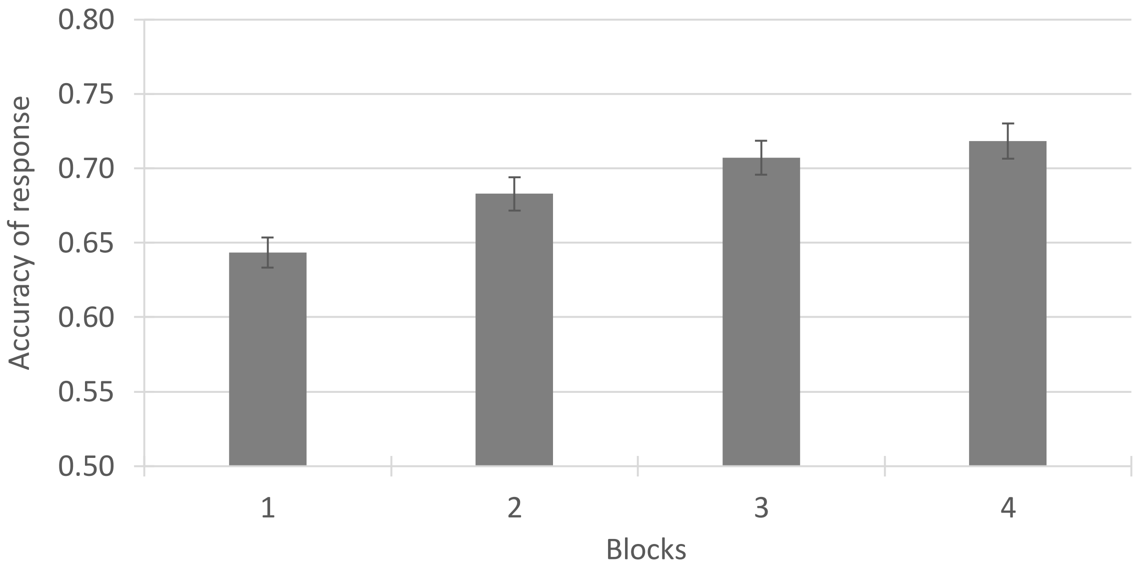


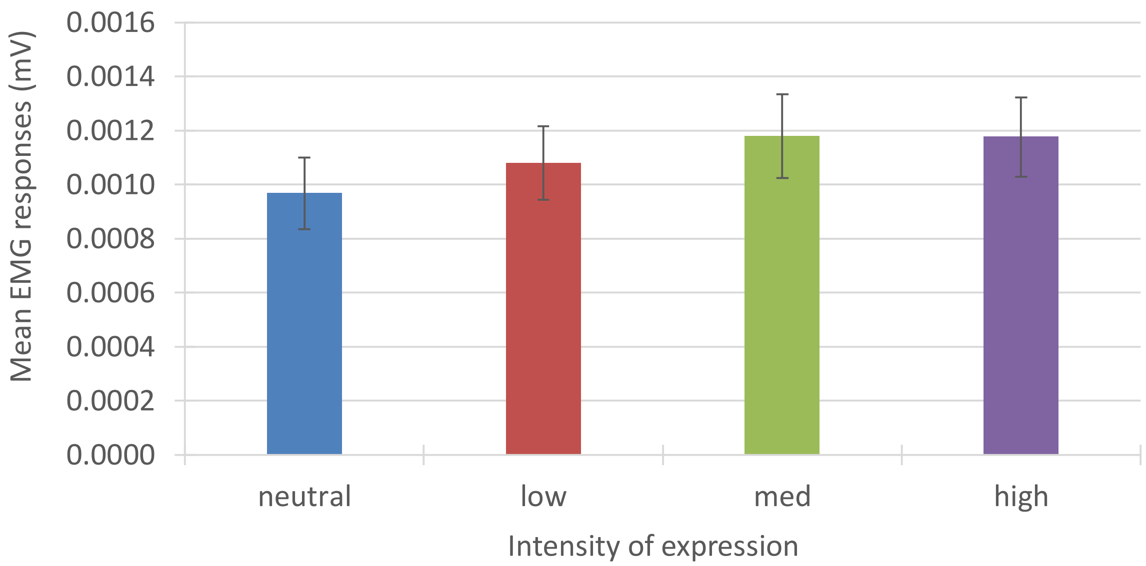
Figure53. Accuracy of response for the facial emotion recognition task divided in four blocks of 90 videos each. Accuracy of responses are expressed in decimal fractions. Error bars represent standard errors of the means.

Conclusion

It was found that accuracy of response indeed changed in that it significantly increased over the first three blocks of the experiment, but not from the third to the last block. The improvements in accuracy of response need to be taken into account for within-subject study designs including experimental manipulations and investigating how these manipulations affect accuracy of response.

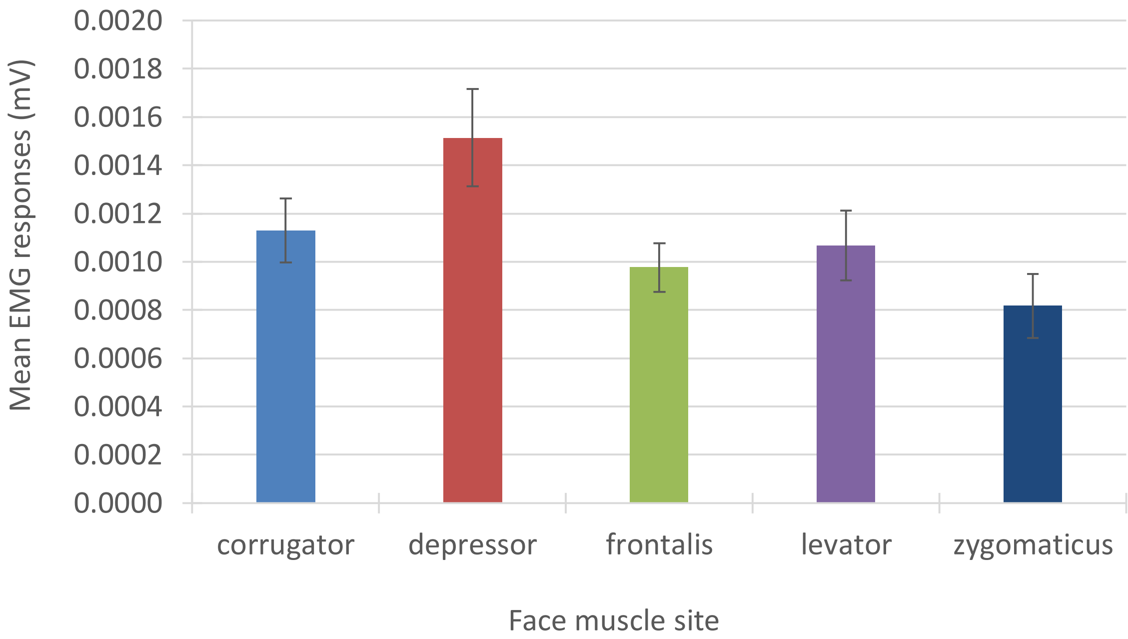
Appendix H. Raw EMG responses for the intensity levels.

The graph displays the EMG responses for each intensity level during facial mimicry based on the untransformed data. Error bars represent the standard errors of the means.



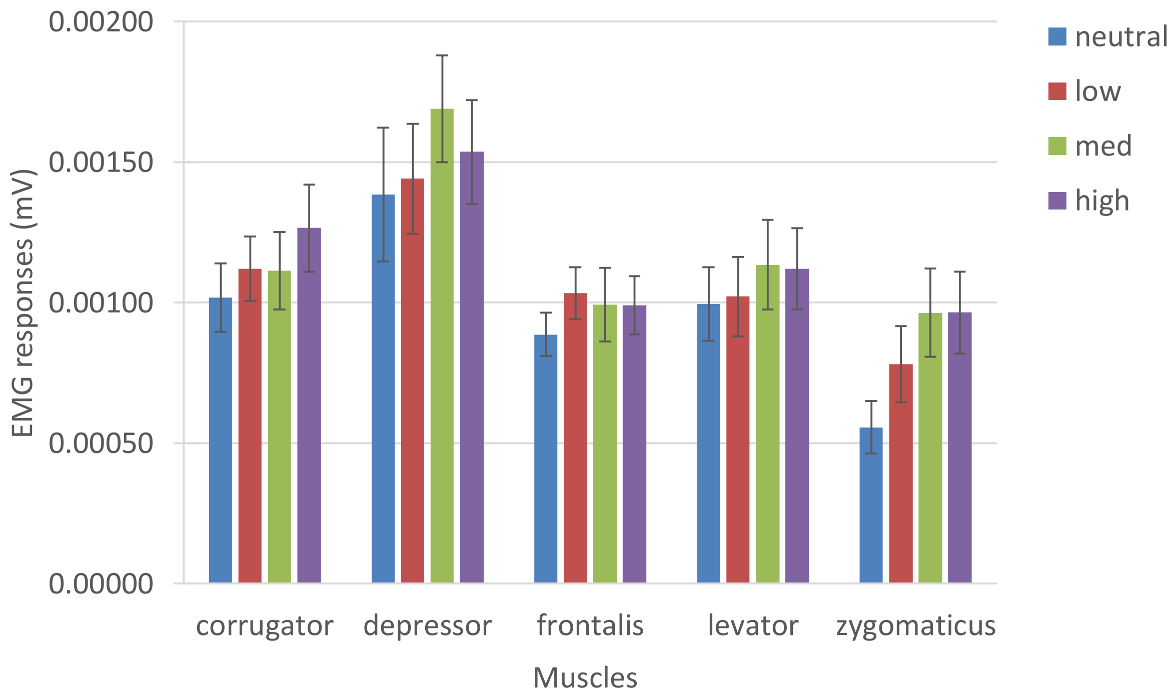
Appendix I. Raw EMG responses for the face muscle sites.

The graph displays the EMG responses for each face muscle site during facial mimicry based on the untransformed data. Error bars represent standard errors of the means.



Appendix J. Raw EMG responses for the face muscle sites by intensity.

The graph displays the EMG responses at each level of intensity of expression for each face muscle site during facial mimicry based on the untransformed data. Error bars represent the standard errors of the means.



REFERENCES

Achaibou, A., Pourtois, G., Schwartz, S., & Vuilleumier, P. (2008). Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia*, *46*(4), 1104–1113. http://doi.org/10.1016/j.neuropsychologia.2007.10.019

Adkins Wood, D. (1960). *Test construction: Development and interpretation of achievement tests*. Columbus: Charles E. Merill Books, Inc.

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62. http://doi.org/10.1177/1534582302001001003

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*(5), 403–410. http://doi.org/10.1111/j.0956-7976.2005.01548.x

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Association.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. E. (2003). Neural correlates of the automatic processing of threat facial signals. *The Journal of Neuroscience*, *23*(13), 5627–33. http://doi.org/23/13/5627 [pii]

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., & Bullmore, E. T. (2007). Differential activation of the amygdala and the "social brain" during fearful face-processing in Asperger Syndrome. *Neuropsychologia*, *45*(1), 2–14. http://doi.org/10.1016/j.neuropsychologia.2006.04.014

Ashwin, C., Chapman, E., Colle, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience*, *1*(3-4), 349–363. http://doi.org/10.1080/17470910601040772

Babchuk, W. A., Hames, R. B., & Thompson, R. A. (1985). Sex differences in the recognition of infant facial expressions of emotion: The primary caretaker hypothesis. *Ethology and Sociobiology*, *6*(2), 89–101. http://doi.org/10.1016/0162-3095(85)90002-0

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*(5), 691–704. http://doi.org/10.1037/a0017088

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161–1179. http://doi.org/10.1037/a0025827

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248–254. http://doi.org/10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*, *21*(1), 37–46. http://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., O’Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning Autism. *Journal of Autism and Developmental Disorders*, *29*(5), 407–418. http://doi.org/10.1023/A:1023035012436

Baron-Cohen, S., Wheelwright, S., Skinner, M., Martin, J., & Clubley, E. (2001). No Title. *Journal of Autism and Developmental Disorders*, *31*(6), 603–603. http://doi.org/10.1023/A:1017455213300

Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005). A first evaluation study of a database of kinetic facial expressions (DaFEx). In *Proceedings of the 7th International Conference on Multimodal Interfaces ICMI 2005, October 04-06, 2005, Trento (Italy)* (pp. 214–221). New York, NY, US: ACM Press.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia*, *44*(1), 110–129. ttp://doi.org/10.1016/j.neuropsychologia.2005.04.002

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*(3), 292–299. http://doi.org/10.1111/j.1469-8986.1985.tb01603.x

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior*, *23*(1), 5–41. http://doi.org/10.1023/A:1021370825283

Blakemore, S.-J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*(2), 260–267. http://doi.org/10.1016/j.neuropsychologia.2004.11.012

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. http://doi.org/10.1016/j.tree.2008.10.008

Bölte, S., & Poustka, F. (2003). The recognition of facial affect in autistic and schizophrenic subjects and their first-degree relatives. *Psychological Medicine*, *33*(5), 907–915. http://doi.org/10.1017/S0033291703007438

Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology*, *11*(3), 263–280. http://doi.org/10.1177/0022022180113003

Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology*, *99*(2), 167–189. http://doi.org/10.1348/000712607X206702

Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion*, *22*(8), 1569–1587. http://doi.org/10.1080/02699930801921156

Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, *38*(1), 22–34. http://doi.org/10.1111/1469-8986.3810022

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. http://doi.org/10.1016/0005-7916(94)90063-9

Brandt, F. S., & Cazzaniga, A. (2007). Hyaluronic Acid Fillers: Restylane and Perlane. *Facial Plastic Surgery Clinics of North America*, *15*(1), 63–76. http://doi.org/10.1016/j.fsc.2006.11.002

Briegel, W. (2006). Neuropsychiatric findings of Möbius sequence - a review. *Clinical Genetics*, *70*(2), 91–97. http://doi.org/10.1111/j.1399-0004.2006.00649.x

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., … Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, *22*(2), 131–152. http://doi.org/10.1068/p220131

Buck, R. (1988). Nonverbal Communication: Spontaneous and Symbolic Aspects. *American Behavioral Scientist*, *31*(3), 341–354. http://doi.org/10.1177/000276488031003006

Buitelaar, J. K., Wees, M. van der, Swaab-Barneveld, H., & Gaag, R. J. van der. (1999). Verbal memory and performance IQ predict theory of mind and emotion recognition ability in children with autistic spectrum disorders and in psychiatric control children. *Journal of Child Psychology and Psychiatry*, *40*(6), 869–881. http://doi.org/10.1111/1469-7610.00505

Bush, L. K., Hess, U., & Wolford, G. (1993). Transformations for within-subject designs: A Monte Carlo investigation. *Psychological Bulletin*, *113*(3), 566–579. http://doi.org/10.1037/0033-2909.113.3.566

Cacioppo, J. T., Uchino, B. N., Crites, S. L., Snydersmith, M. A., Smith, G., Berntson, G. G., & Lang, P. J. (1992). Relationship between facial expressiveness and sympathetic activation in emotion: A critical review, with emphasis on modeling underlying mechanisms and individual differences. *Journal of Personality and Social Psychology*. US: American Psychological Association. http://doi.org/10.1037/0022-3514.62.1.110

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 527–551. http://doi.org/10.1037/0096-1523.26.2.527

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., & Nummenmaa, L. (2014). Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*(4), 549–567. http://doi.org/10.1007/s10919-014-0191-3

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. http://doi.org/10.3758/BRM.40.1.109

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, *137*(3), 471–494. http://doi.org/10.1037/a0012771

Calvo, M. G., & Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: The role of visual saliency. *Vision Research*, *51*(15), 1751–1759. http://doi.org/10.1016/j.visres.2011.06.001

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of Simple and Complex Emotions in Non-retarded Children with Autism. *Journal of Child Psychology and Psychiatry*, *33*(7), 1169–1182. http://doi.org/10.1111/j.1469-7610.1992.tb00936.x

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*. US: American Psychological Association. http://doi.org/10.1037/0022-3514.76.6.893

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*(1), 285–308. http://doi.org/10.1146/annurev-psych-113011-143754

Chartrand, T. L., & van Baaren, R. (2009). Human Mimicry. In *Advances in experimental social psychology, Vol 41* (pp. 219–274). San Diego, CA, US: Elsevier Academic Press. http://doi.org/10.1016/S0065-2601(08)00405-X

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion*, *7*(4), 802–811. http://doi.org/10.1037/1528-3542.7.4.802

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*, *48*(1), 220–225. http://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2009.09.007

Conson, M., Ponari, M., Monteforte, E., Ricciato, G., Sarà, M., Grossi, D., & Trojano, L. (2013). Explicit recognition of emotional facial expressions is shaped by expertise: evidence from professional actors. *Frontiers in Psychology*, *4*. http://doi.org/10.3389/fpsyg.2013.00382

Corp., I. (2012). IBM SPSS Statistics for Windows, Version 21.0. Armonk: IBM Corp.

Corp., I. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk: IBM Corp.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In M. J. W. J. T. Townsend (Ed.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 347–395). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Cram, J. R. (1998). *Cram's introduction to surface electromyography* (2nd ed.). Sudbury, MA, USA: Jones and Bartlett Publishers.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando: Wadsworth Publishing Co Inc.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*(3), 497–507. http://doi.org/10.1016/j.neuron.2005.10.018

Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York, NY, US: Grosset/Putnam.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness.* Fort Worth, TX, US: Harcourt College Publishers.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), 28–30. http://doi.org/10.1038/nn1611

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, England: John Murray. http://doi.org/10.1037/10001-000

Davis, J. I., Senghas, A., Brandt, F., & Ochsner, K. N. (2010). The effects of BOTOX injections on emotional experience. *Emotion*, *10*(3), 433–440. http://doi.org/10.1037/a0018690

Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Development*, *73*(3), 700–17.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the Nature of Face Processing Impairment in Autism: Insights From Behavioral and Electrophysiological Studies. *Developmental Neuropsychology*, *27*(3), 403–424. http://doi.org/10.1207/s15326942dn2703_6

De Sonneville, L. M. J., Verschoor, C. A., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., & Vranken, M. (2002). Facial identity and facial emotions: speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, *24*(2), 200–213. http://doi.org/10.1076/jcen.24.2.200.989

Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, *5*(1), 126–134. http://doi.org/10.1016/j.rasd.2010.03.002

Derntl, B., Habel, U., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2009). General and specific responsiveness of the amygdala during explicit emotion recognition in females and males. *BMC Neuroscience*, *10*(1), 91. http://doi.org/10.1186/1471-2202-10-91

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. http://doi.org/10.1007/BF00230027

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, *19*(6), 643–647. http://doi.org/10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, *39*(1), 39–45. http://doi.org/10.1111/1467-9450.00054

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*(1), 86–89. http://doi.org/10.1111/1467-9280.00221

Doi, H., Fujisawa, T. X., Kanai, C., Ohta, H., Yokoi, H., Iwanami, A., … Shinohara, K. (2013). Recognition of facial expressions and prosodic cues with graded emotional intensities in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *43*(9), 2099–2113. http://doi.org/10.1007/s10803-013-1760-8

Donges, U.-S., Kersting, A., & Suslow, T. (2012). Women's Greater Ability to Perceive Happy Facial Emotion Automatically: Gender Differences in Affective Priming. *PLoS ONE*, *7*(7), e41745. http://doi.org/10.1371/journal.pone.0041745

Duchenne de Bologne, G. B. (1862). *The mechanism of human facial expression (RA Cuthbertson, Trans.)*. Paris: Jules Renard.

Dyck, M. J. (2012). The ability to understand the experience of other people: Development and validation of the Emotion Recognition Scales. *Australian Psychologist*, *47*(1), 49–57. http://doi.org/10.1111/j.1742-9544.2011.00047.x

Eack, S. M., Mazefsky, C. A., & Minshew, N. J. (2015). Misinterpretation of facial expressions of emotion in verbal adults with autism spectrum disorder. *Autism*, *19*(3), 308–315. http://doi.org/10.1177/1362361314520755

Eigsti, I.-M. (2013). A review of embodiment in autism spectrum disorders. *Frontiers in Psychology*, *4*. http://doi.org/10.3389/fpsyg.2013.00224

Ekman, P. (1987). Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social …*, *53*, 712–717.

Ekman, P. (1992a). An argument for basic emotions. *Cognition & Emotion*, *6*(3), 169–200. http://doi.org/10.1080/02699939208411068

Ekman, P. (1992b). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. http://doi.org/10.1037/0033-295X.99.3.550

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. http://doi.org/10.1037/0003-066X.48.4.384

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*(2), 268–287. http://doi.org/10.1037/0033-2909.115.2.268

Ekman, P. (2005). Facial Expressions. In T. D. M. J. Power (Ed.), *Handbook of Cognition and Emotion* (pp. 301–320). Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/0470013494.ch16

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. http://doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. http://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Oxford, England: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–258. http://doi.org/10.1007/BF00987191

Ekman, P., Friesen, W. V, & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behaviour? In, P. Ekman. In P. Ekman (Ed.), *Emotion in the Human Face* (2nd ed., pp. 39–55). Cambridge: Cambridge University Press.

Ekman, P., Friesen, W. V, & Hager, J. C. (2002). *Facial action coding system*. Salt Lake City, UT: Research Nexus.

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect*. Palo Alto: Consulting Psychologists Press.

Ekman, P., & Friesen, W. (1978). *Facial Action Coding System: A technique for the measurement of facial movements*. Palo Alto: Consulting Psychologists Press.

Ekman, P., & Friesen, W. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, *10*(2), 159–168. http://doi.org/10.1007/BF00992253

Ekman, P., Levenson, R., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210. http://doi.org/10.1126/science.6612338

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, *30*(1), 527–554. http://doi.org/10.1146/annurev.ps.30.020179.002523

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88. http://doi.org/10.1126/science.164.3875.86

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. http://doi.org/10.1037/0033-2909.128.2.203

Enticott, P. G., Johnston, P. J., Herring, S. E., Hoy, K. E., & Fitzgerald, P. B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia*, *46*(11), 2851–2854. http://doi.org/10.1016/j.neuropsychologia.2008.04.022

Field, A. (2009). *Discovering statistics using SPSS*. London, England: SAGE.

Fogassi, L., & Ferrari, P. F. (2004). Mirror neurons, gestures and language evolution. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, *5*(3), 345–363. http://doi.org/10.1075/is.5.3.03fog

Fombonne, E. (2005). The changing epidemiology of autism. *Journal of Applied Research in Intellectual Disabilities*, *18*(4), 281–294. http://doi.org/10.1111/j.1468-3148.2005.00266.x

Foroni, F., & Semin, G. R. (2011). When does mimicry affect evaluative judgment? *Emotion*, *11*(3), 687–690. http://doi.org/10.1037/a0023163

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The Neural Basis of Contextual Influences on Face Categorization. *Cerebral Cortex*, *25*(2), 415–422. http://doi.org/10.1093/cercor/bht238

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS ONE*, *8*(1), e53286. http://doi.org/10.1371/journal.pone.0053286

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, *60*(2), 229–240. http://doi.org/10.1037/0022-3514.60.2.229

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA, US: Academic Press.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for Human Electromyographic Research. *Psychophysiology*, *23*(5), 567–589. http://doi.org/10.1111/j.1469-8986.1986.tb00676.x

Frijda, N. H. (1953). The understanding of facial expression of emotion. *Acta Psychologica*, *9*, 294–362. http://doi.org/10.1016/0001-6918(53)90021-3

Frith, U. (2003). *Autism: explaining the enigma (2nd ed.).* Malden: Blackwell Publishing.

Fujimura, T., Matsuda, Y.-T., Katahira, K., Okada, M., & Okanoya, K. (2012). Categorical and dimensional perceptions in decoding emotional facial expressions. *Cognition & Emotion*, *26*, 587–601. http://doi.org/10.1080/02699931.2011.595391

Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, *36*(4), 171–180. http://doi.org/10.1159/000072786

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. http://doi.org/10.1093/brain/119.2.593

Gauthier, I., & Tarr, M. (1997). Becoming a “greeble” expert: exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682. http://doi.org/10.1016/S0042-6989(96)00286-6

Gioia, J. V, & Brosgole, L. (1988). Visual and auditory affect recognition in singly diagnosed mentally retarded patients, mentally retarded patients with autism and normal young children. *International Journal of Neuroscience*, *43*(3-4), 149–163. http://doi.org/10.3109/00207458808986164

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213. http://doi.org/10.1016/j.cognition.2004.01.005

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: what autism teaches us about face processing. *Developmental Psychobiology*, *40*(3), 213–25. http://doi.org/10.1002/dev.10028

Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition*, *54*(3), 248–250. http://doi.org/10.1016/j.bandc.2004.02.029

Gross, T. F. (2004). The perception of four basic emotions in human and nonhuman faces by children with autism and other developmental disabilities. *Journal of Abnormal Child Psychology*, *32*(5), 469–480. http://doi.org/10.1023/B:JACP.0000037777.17698.01

Gueguen, N., Jacob, C., & Martin, A. (2009). Mimicry in social interaction: Its effect on human judgment and behavior. *European Journal of Social Sciences*, *8*(2), 253–259.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*(4), 845–857. http://doi.org/10.1037/0033-2909.85.4.845

Hall, J. A. (1990). *Nonverbal sex differences: Accuracy of communication and expressive style.* Johns Hopkins University Press.

Hall, J., Hutton, S., & Morgan, M. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition & Emotion*. http://doi.org/10.1080/02699930902906882

Hall, J., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201–206. http://doi.org/10.1037/1528-3542.4.2.201

Hampson, E., Van Anders, S., & Mullin, L. (2006). A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis. *Evolution and Human Behavior*, *27*(6), 401–416. http://doi.org/10.1016/j.evolhumbehav.2006.05.002

Handa, R. J., & McGivern, R. F. (2015). Steroid Hormones, Receptors, and Perceptual and Cognitive Sex Differences in the Visual System. *Current Eye Research*, *40*(2), 110–127. http://doi.org/10.3109/02713683.2014.952826

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. http://doi.org/10.1007/s10803-005-0039-0

Hara, & Kobayashi. (1995). *Experimental research on dynamic recognition system for human facial expressions: Development of recognition and generation system for facial expression.* In Psychological and informatics research on Kansei information processing. Tokyo: MEXT.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. http://doi.org/10.1007/s11065-010-9138-6

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal on Mental Retardation*, *104*(3), 270–278. http://doi.org/10.1352/0895-8017(1999)104<0270:ROFEEF>2.0.CO;2

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: a naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, *30*(3), 225–236. http://doi.org/10.1023/A:1005544518785

Heerdink, M. W., van Kleef, G. A., Homan, A. C., & Fischer, A. H. (2015). Emotional expressions as social signals of rejection and acceptance: Evidence from the Affect Misattribution Paradigm. *Journal of Experimental Social Psychology*, *56*, 60–68. http://doi.org/10.1016/j.jesp.2014.09.004

Hennenlotter, A., Dresel, C., Castrop, F., Ceballos-Baumann, A. O., Wohlschlager, A. M., & Haslinger, B. (2009). The link between facial feedback and neural activity within central circuitries of emotion--new insights from Botulinum toxin-induced denervation of frown muscles. *Cerebral Cortex*, *19*(3), 537–542. http://doi.org/10.1093/cercor/bhn104

Hess, U., Adams, R. B., & Kleck, R. E. (2009). The face is not an empty canvas: how facial expressions interact with facial appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3497–3504. http://doi.org/10.1098/rstb.2009.0165

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, *40*(2), 129–141. http://doi.org/10.1016/S0167-8760(00)00161-6

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*(4), 241–257. http://doi.org/10.1023/A:1024952730333

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. http://doi.org/10.1177/1088868312472607

Hess, U., & Fischer, A. (2014). Emotional mimicry: why and when we mimic emotions. *Social and Personality Psychology Compass*, *8*(2), 45–57. http://doi.org/10.1111/spc3.12083

Hess, U., Kappas, A., McHugo, G. J., Lanzetta, J. T., & Kleck, R. E. (1992). The facilitative effect of facial expression on the self-generation of emotion. *International Journal of Psychophysiology*, *12*(3), 251–265. http://doi.org/10.1016/0167-8760(92)90064-I

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*(5), 369–385. http://doi.org/10.1002/ejsp.2420200502

Hess, U., Philippot, P., & Blairy, S. (1998). Facial reactions to emotional facial expressions: affect or cognition? *Cognition & Emotion*, *12*(4), 509–531. http://doi.org/10.1080/026999398379547

Hillier, A., & Allinson, L. (2002). Understanding Embarrassment Among Those with Autism: Breaking Down the Complex Emotion of Embarrassment Among Those with Autism. *Journal of Autism and Developmental Disorders*, *32*(6), 583–592. http://doi.org/10.1023/A:1021259115185

Hinsz, V. B., & Tomhave, J. A. (1991). Smile and (half) the world smiles with you, frown and you frown alone. *Personality and Social Psychology Bulletin*, *17*(5), 586–592. http://doi.org/10.1177/0146167291175014

Hoekstra, R. A., Vinkhuyzen, A. A. E., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., … van der Sluis, S. (2011). The construction and validation of an abridged version of the Autism-Spectrum Quotient (AQ-Short). *Journal of Autism and Developmental Disorders*, *41*(5), 589–596. http://doi.org/10.1007/s10803-010-1073-0

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278–283. http://doi.org/10.1016/j.actpsy.2010.07.012

Hoffmann, H., Traue, H. C., Bachmayr, F., & Kessler, H. (2010). Perceived realism of dynamic facial expressions of emotion: Optimal durations for the presentation of emotional onsets and offsets. *Cognition & Emotion*, *24*(8), 1369–1376. http://doi.org/10.1080/02699930903417855

Horstmann, G. (2003). What do facial expressions convey: feeling states, behavioral intentions, or action requests? *Emotion*, *3*(2), 150–66.

Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. *Neuroreport*, *11*(13), 2931–2935.

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, *45*(4), 685–95. http://doi.org/10.1016/j.neuropsychologia.2006.08.003

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951. http://doi.org/10.1038/nrn2024

Izard, C. E. (1968). Cross-cultural research findings on development in recognition of facial behavior. In *symposium" Universality of the Emotions" of the American Psychological Association*.

Izard, C. E. (1969). The emotions and emotion constructs in personality and culture research. *Handbook of Modern Personality Theory*, 496–510.

Izard, C. E. (1971). *The face of emotion*. East Norwalk, CT, US: Appleton-Century-Crofts.

Izard, C. E. (1977). *Human emotions*. *NY: Plenum*. New York, NY, US: Plenum Publishing Corp.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2), 187–192. http://doi.org/http://dx.doi.org/10.1016/j.cub.2013.11.064

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., … Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *52*(3), 275–285. http://doi.org/10.1111/j.1469-7610.2010.02328.x

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2013). Dynamic properties influence the perception of facial expressions. *Perception*, *42*(11), 1266–1278. http://doi.org/10.1068/p3131n

Katsikitis, M. (1997). The classification of facial expressions of emotion: a multidimensional-scaling approach. *Perception*, *26*(5), 613–626. http://doi.org/10.1068/p260613

Kaulard, K., Cunningham, D. W., Bulthoff, H. H., & Wallraven, C. (2012). The MPI facial expression database--a validated database of emotional and conversational facial expressions. *Plos One*, *7*(3), e32321. http://doi.org/10.1371/journal.pone.0032321

Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Leahy, P., … Wirsing, R. (1981). Culture and the perception of social dominance from facial expression. *Journal of Personality and Social Psychology*, *40*(4), 615–626. http://doi.org/10.1037/0022-3514.40.4.615

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, *13*(5), 505–521. http://doi.org/10.1080/026999399379168

Kennedy, D. P., & Adolphs, R. (2012). Perception of emotions from facial expressions in high-functioning adults with autism. *Neuropsychologia*, *50*(14), 3313–3319. http://doi.org/10.1016/j.neuropsychologia.2012.09.038

Kestenbaum, R., & Nelson, C. A. (1992). Neural and behavioral correlates of emotion recognition in children and adults. *Journal of Experimental Child Psychology*, *54*(1), 1–18. http://doi.org/10.1016/0022-0965(92)90014-W

Khan, R. A., Meyer, A., Konik, H., & Bouakaz, S. (2012). Human vision inspired framework for facial expressions recognition. In *2012 19th IEEE International Conference on Image Processing* (pp. 2593–2596). IEEE. http://doi.org/10.1109/ICIP.2012.6467429

Kirouac, G., & Dore, F. Y. (1985). Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior*, *9*(1), 3–7. http://doi.org/10.1007/BF00987555

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H. R. (2010). Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. *Journal of Neuroscience*, *30*(37), 12281–12287. http://doi.org/10.1523/JNEUROSCI.0688-10.2010

Koerner, N., & Antony, M. M. (2010). Special series on disgust and phobic avoidance: a commentary. *International Journal of Cognitive Therapy*, *3*(1), 52–63. http://doi.org/10.1521/ijct.2010.3.1.52

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221. http://doi.org/10.1016/j.neuropsychologia.2011.12.022

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*(1), 41–46. http://doi.org/10.1177/1754073912451349

Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., & Wilhelm, O. (2014). Facial EMG Responses to Emotional Expressions Are Related to Emotion Perception Ability. *PLoS ONE*, *9*(1), e84053. http://doi.org/10.1371/journal.pone.0084053

Kurita, H., Koyama, T., & Osada, H. (2005). Autism-Spectrum Quotient-Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry and Clinical Neurosciences*, *59*(4), 490–6. http://doi.org/10.1111/j.1440-1819.2005.01403.x

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162. http://doi.org/10.1023/A:1025389814290

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–85.

Law Smith, M. J., Montagne, B., Perrett, D. I., Gill, M., & Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning Autism using dynamic stimuli of varying intensities. *Neuropsychologia*, *48*(9), 2777–2781. http://doi.org/10.1016/j.neuropsychologia.2010.03.008

Lee, N. C., Krabbendam, L., White, T. P., Meeter, M., Banaschewski, T., Barker, G. J., … Shergill, S. S. (2013). Do you see what I see? Sex differences in the discrimination of facial emotions during adolescence. *Emotion*, *13*(6), 1030–1040. http://doi.org/10.1037/a0033560

Lee, T.-W., Josephs, O., Dolan, R. J., & Critchley, H. D. (2006). Imitating expressions: emotion-specific neural substrates in facial mimicry. *Social Cognitive and Affective Neuroscience*, *1*(2), 122–35. http://doi.org/10.1093/scan/nsl012

Leppänen, J. M., & Hietanen, J. K. (2007). Is there more in a happy face than just a big smile? *Visual Cognition*, *15*(4), 468–490. http://doi.org/10.1080/13506280600765333

Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience*, *10*(1), 37–47. http://doi.org/10.1038/nrn2554

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, *6*(4), 315–324. http://doi.org/10.1111/j.2044-835X.1988.tb01104.x

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage*, *21*(2), 601–607. http://doi.org/10.1016/j.neuroimage.2003.09.038

Levenson, R. W., Carstensen, L. L., Friesen, W. V, & Ekman, P. (1991). Emotion, physiology, and expression in old age. *Psychology and Aging*, *6*(1), 28–35. http://doi.org/10.1037/0882-7974.6.1.28

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity. *Psychophysiology*, *27*(4), 363–384. http://doi.org/10.1111/j.1469-8986.1990.tb02330.x

Likowski, K. U., Mühlberger, A., Gerdes, A. B. M., Wieser, M. J., Pauli, P., & Weyers, P. (2012). Facial mimicry and the mirror neuron system: simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Frontiers in Human Neuroscience*, *6*. http://doi.org/10.3389/fnhum.2012.00214

Limbrecht-Ecklundt, K., Scheck, A., Jerg-Bretzke, L., Walter, S., Hoffmann, H., & Traue, H. C. (2013). The effect of forced choice on facial emotion recognition: a comparison to open verbal classification of emotion labels. *GMS Psycho-Social-Medicine*, *10*, Doc04. http://doi.org/10.3205/psm000094

Lipps, T. (1907). *Das Wissen von fremden Ichen*. *Psychologische Untersuchungen* (Vol. 1). Leipzig: Engelmann.

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.01171

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule: ADOS-2*. LA: Western Psychological Services Los Angeles, CA.

Losh, M., & Capps, L. (2006). Understanding of emotional experience in autism: Insights from the personal accounts of high-functioning children with autism. *Developmental Psychology*, *42*(5), 809–818. http://doi.org/10.1037/0012-1649.42.5.809

Lozier, L. M., Vanmeter, J. W., & Marsh, A. A. (2014). Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. *Development and Psychopathology*, *26*(4pt1), 933–945. http://doi.org/10.1017/S0954579414000479

Lundqvist, L.-O. (1995). Facial EMG reactions to facial expressions: A case of facial emotional contagion? *Scandinavian Journal of Psychology*, *36*(2), 130–141. http://doi.org/10.1111/j.1467-9450.1995.tb00974.x

Lundqvist, L.-O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*, *9*(3), 203–211.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion*, *5*(1), 119–124. http://doi.org/10.1037/1528-3542.5.1.119

Marsh, A. A., Kozak, M. N., & Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion*, *7*(2), 239–251. http://doi.org/10.1037/1528-3542.7.2.239

Matsumoto, D., & Ekman, P. (1988). Japanese and Caucasian facial expressions of emotion and neutral faces (JACFEE and JACNeuF). *Human Interaction Laboratory, University of California, San Francisco*, *401*.

Matsumoto, D., & Ekman, P. (2004). The relationship among expressions, labels, and descriptions of contempt. *Journal of Personality and Social Psychology*, *87*(4), 529–540. http://doi.org/10.1037/0022-3514.87.4.529

Matsumoto, D., & Hwang, H. (2011). Reading facial expressions of emotion. *Psychological Science Agenda, Science Brief*, *Vol 25*(5). Retrieved from http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx

Matsumoto, D., & Hwang, H. C. (2014). Judgments of subtle facial expressions of emotion. *Emotion*, *14*(2), 349–357. http://doi.org/10.1037/a0035237

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions (3rd ed.)* (pp. 211–234). New York, NY, US: Guilford Press.

Mazefsky, C. A., & Oswald, D. P. (2007). Emotion perception in Asperger's syndrome and high-functioning autism: the tmportance of diagnostic criteria and cue intensity. *Journal of Autism and Developmental Disorders*, *37*(6), 1086–1095. http://doi.org/10.1007/s10803-006-0251-6

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*(2), 211–224. http://doi.org/10.1177/0022002186017002005

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424–453. http://doi.org/10.1037/0033-2909.126.3.424

McIntosh, D. N. (2006). Spontaneous facial mimicry, liking and emotional contagion. *Polish Psychological Bulletin*, *37*(1), 31–42.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science*, *9*(3), 295–302. http://doi.org/10.1111/j.1467-7687.2006.00492.x

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8–15. http://doi.org/10.1016/j.tics.2006.11.002

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*, *36*(1), 341–349. http://doi.org/10.1016/j.neubiorev.2011.07.004

Mondloch, C. J., Grand, R. Le, & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*(5), 553–566. http://doi.org/10.1068/p3339

Montagne, B., Kessels, R. P. C., De Haan, E. H. F., & Perrett, D. I. (2007). The Emotion Recognition Task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills*, *104*(2), 589–598. http://doi.org/10.2466/pms.104.2.589-598

Montgomery, K. J., & Haxby, J. V. (2008). Mirror neuron system differentially activated by facial expressions and social hand gestures: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*(10), 1866–1877. http://doi.org/10.1162/jocn.2008.20127

Morton, A. (1980). *Frames of mind: Constraints on the common-sense conception of the mental*. Oxford, England: Oxford University Press.

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, *52*(1), 1–22. http://doi.org/10.1080/10570318809389622

Mufson, L., & Nowicki, S. (1991). Factors affecting the accuracy of facial affect recognition. *The Journal of Social Psychology*, *131*(6), 815–822.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, *20*(8), 750–756. http://doi.org/10.1016/j.cub.2010.02.045

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678. http://doi.org/10.1177/1948550611406138

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002–1005. http://doi.org/10.1126/science.1136930

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion*, *15*, 853–864. http://doi.org/10.1080/02699930143000194

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, Å. H. (2000). Emotional state and the detection of change in facial expression of emotion. *European Journal of Social Psychology*, *30*(2), 211–222. http://doi.org/10.1002/(SICI)1099-0992(200003/04)30:2<211::AID-EJSP988>3.0.CO;2-3

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*(06), 417–433. http://doi.org/10.1017/S0140525X10000865

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, *18*(1), 9–35. http://doi.org/10.1007/BF02169077

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, *111*(2), 646–651. http://doi.org/10.1073/pnas.1321664111

Nuske, H. J., Vivanti, G., & Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognition & Emotion*, *27*(6), 1042–1061. http://doi.org/10.1080/02699931.2012.762900

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(5), 812–816. http://doi.org/10.1109/TPAMI.2005.90

Oatley, K., & Johnson-laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition & Emotion*, *1*(1), 29–50. http://doi.org/10.1080/02699938708408362

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social Neuroscience*, *2*(3-4), 167–178. http://doi.org/10.1080/17470910701391943

Ogawa, T., & Suzuki, N. (1999). Response differentiation to facial expression of emotion as increasing exposure duration. *Perceptual and Motor Skills*, *89*(2), 557–563.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. http://doi.org/10.1037/0033-295X.97.3.315

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–261. http://doi.org/10.1023/A:1016374617369

Perkins, T., Stokes, M., McGillivray, J., & Bittar, R. (2010). Mirror neuron dysfunction in autism spectrum disorders. *Journal of Clinical Neuroscience : Official Journal of the Neurosurgical Society of Australasia*, *17*(10), 1239–1243. http://doi.org/10.1016/j.jocn.2010.01.026

Piggot, J., Kwon, H., Mobbs, D., Blasey, C., Lotspeich, L., Menon, V., … Reiss, A. L. (2004). Emotional attribution in high-functioning individuals with autistic spectrum disorder: A functional imaging study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(4), 473–480. http://doi.org/10.1097/00004583-200404000-00014

Plutchik, R. (1980). *A general psychoevolutionary theory of emotion*. *Theories of emotion* (Vol. 1). New York, NY: Academic Press.

Pollick, F. E., Hill, H., Calder, A., & Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception*, *32*(7), 813–826. http://doi.org/10.1068/p3319

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., & Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion*, *12*(6), 1398–1403. http://doi.org/10.1037/a0028588

Psychology Software Tools, I. (2012). [E-Prime 2.0]. Pittsburgh, PA. Retrieved from http://www.pstnet.com

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition*, *54*(3), 179–185. http://doi.org/10.1016/j.bandc.2004.01.002

Ramachandran, V. S., & Oberman, L. M. (2008). Broken mirrors: A theory of autism. In *The Jossey-Bass reader on the brain and learning* (pp. 435–443). San Francisco, CA, US: Jossey-Bass.

Recio, G., Schacht, A., & Sommer, W. (2013). Classification of dynamic facial expressions of emotion presented briefly. *Cognition and Emotion*, *27*(8), 1486–1494. http://doi.org/10.1080/02699931.2013.794128

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. *Journal of Autism and Developmental Disorders*, *20*(4), 555–567. http://doi.org/10.1007/BF02216060

Rhodes, M. G. (2009). Age estimation of faces: a review. *Applied Cognitive Psychology*, *23*(1), 1–12. http://doi.org/10.1002/acp.1442

Rives Bogart, K., & Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience*, *5*(2), 241–251. http://doi.org/10.1080/17470910903395692

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–41.

Robertson, C. E., Thomas, C., Kravitz, D. J., Wallace, G. L., Baron-Cohen, S., Martin, A., & Baker, C. I. (2014). Global motion perception deficits in autism are reflected as early as primary visual cortex. *Brain*, *137*(9), 2588–2599. http://doi.org/10.1093/brain/awu189

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion*, *8*(3), 201–229. http://doi.org/10.1080/02699939408408938

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal cues: The PONS test*. Baltimore: The Johns Hopkins University Press.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139–148. http://doi.org/10.1007/BF00986931

Roy-Charland, A., Perron, M., Beaudry, O., & Eady, K. (2014). Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Cognition and Emotion*, *28*(7), 1214–1222. http://doi.org/10.1080/02699931.2013.878687

Rozin, P., Taylor, C., Ross, L., Bennett, G., & Hejmadi, A. (2005). General and specific abilities to recognise negative emotions, especially disgust, as portrayed in the face and the body. *Cognition & Emotion*, *19*(3), 397–412. http://doi.org/10.1080/02699930441000166

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, *32*(4), 863–881. http://doi.org/10.1016/j.neubiorev.2008.01.001

Rump, K. M., Giovannelli, J. L., Minshew, N. J., & Strauss, M. S. (2009). The development of emotion recognition in individuals with autism. *Child Development*, *80*(5), 1434–1447. http://doi.org/10.1111/j.1467-8624.2009.01343.x

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41–51. http://doi.org/10.1007/BF00995206

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–41.

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire.* LA: Western Psychological Services.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Subgrouping siblings of people with autism: Identifying the broader autism phenotype. *Autism Research*. http://doi.org/10.1002/aur.1544

Sackeim, H. A., Gur, R. C., & Saucy, M. C. (1978). Emotions are expressed more intensely on the left side of the face. *Science*, *202*(4366), 434–436.

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology*, *70*(1), 70–74. http://doi.org/10.1016/j.ijpsycho.2008.06.001

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, *20*(1), 81–91. http://doi.org/10.1016/j.cogbrainres.2004.01.008

Sato, W., & Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cognition & Emotion*, *18*(5), 701–710. http://doi.org/10.1080/02699930341000176

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the Emotion Recognition Index. *Journal of Nonverbal Behavior*, *35*(4), 305–326. http://doi.org/10.1007/s10919-011-0115-4

Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience*, *3*(1), 37–50. http://doi.org/10.1080/17470910701563228

Schmid, P., & Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motivation and Emotion*, *34*(3), 288–292. http://doi.org/10.1007/s11031-010-9170-0

Schneider, K. G., Hempel, R. J., & Lynch, T. R. (2013). That "poker face" just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion. *Emotion*, *13*(5), 852–866. http://doi.org/10.1037/a0032847

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, *42*(1), 393–403. http://doi.org/10.1016/j.neuroimage.2008.04.180

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, *135*(1-2), 55–64. http://doi.org/10.1016/j.pain.2007.05.008

Simpson, J. A., & Weiner, E. S. C. (1989). *The Oxford english dictionary* (Vol. 2). Oxford, England: Clarendon Press.

Sinzig, J., Morsch, D., & Lehmkuhl, G. (2008). Do hyperactivity, impulsivity and inattention have an impact on the ability of facial affect recognition in children with autism and ADHD? *European Child & Adolescent Psychiatry*, *17*(2), 63–72. http://doi.org/10.1007/s00787-007-0637-9

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*(5), 929–939. http://doi.org/10.1007/s10803-006-0232-9

Stein, N. L., Hernandez, M. W., & Trabasso, T. (2008). Advances in modeling emotion and thought: The importance of developmental online, and multilevel analyses. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions (3rd ed.)* (pp. 574–586). New York, NY, US: Guilford Press.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768–777. http://doi.org/10.1037/0022-3514.54.5.768

Suzuki, A., Hoshino, T., & Shigemasu, K. (2006). Measuring individual differences in sensitivities to basic emotions in faces. *Cognition*, *99*(3), 327–353. http://doi.org/http://dx.doi.org/10.1016/j.cognition.2005.04.003

Takehara, T., & Suzuki, N. (1997). Morphed images of basic emotional expressions: ratings on Russell's bipolar field. *Perceptual and Motor Skills*, *85*(3), 1003–1010. http://doi.org/10.2466/pms.1997.85.3.1003

Takehara, T., & Suzuki, N. (2001). Robustness of the two-dimensional structure of recognition of facial expression: evidence under different intensities of emotionality. *Perceptual and Motor Skills*, *93*(3), 739–753. http://doi.org/10.2466/pms.2001.93.3.739

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*(6), 657–665. http://doi.org/10.1080/13506280500410949

Thode, H. C. (2002). *Testing for normality.* New York, NY, US: Marcel Dekker. http://doi.org/10.1201/9780203910894

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, *10*(5), 547–558. http://doi.org/10.1111/j.1467-7687.2007.00614.x

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education (5th ed.).* New York, NY, England: Macmillan Publishing Co, Inc.

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. I. The positive affects*. Oxford, England: Springer.

Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (Vol. 163, p. 195). New York, NY, US: Psychology Press.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. http://doi.org/10.1016/j.psychres.2008.05.006

Tracy, J. L., & Robins, R. W. (2007). Emerging insights into the nature and function of pride. *Current Directions in Psychological Science*, *16*(3), 147–150. http://doi.org/10.1111/j.1467-8721.2007.00493.x

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, *8*(1), 81–95. http://doi.org/10.1037/1528-3542.8.1.81

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Research*, *1284*, 100–115. http://doi.org/10.1016/j.brainres.2009.05.075

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. http://doi.org/10.1007/s10803-012-1695-5

Uvnäs-Moberg, K. (1998). Oxytocin may mediate the benefits of positive social interaction and emotions. *Psychoneuroendocrinology*, *23*(8), 819–835. http://doi.org/10.1016/S0306-4530(98)00056-0

Vallbo, A. B. (1974). Human muscle spindle discharge during isometric voluntary contractions. Amplitude relations between spindle frequency and torque. *Acta Physiologica Scandinavica*, *90*(2), 319–336. http://doi.org/10.1111/j.1748-1716.1974.tb05594.x van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). Facial expressions: What the mirror neuron system can and cannot tell us. *Social Neuroscience*, *2*(3-4), 179–222. http://doi.org/10.1080/17470910701376878

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. http://doi.org/10.1037/a0023853

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885. http://doi.org/10.1162/jocn.2009.21366

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. http://doi.org/10.1007/BF00987006

Wagner, H. L. (2000). The accessibility of the term “contempt” and the meaning of the unilateral lip curl. *Cognition and Emotion*, *14*(5), 689–710. http://doi.org/10.1080/02699930050117675

Wallace, G. L., Case, L. K., Harms, M. B., Silvers, J. A., Kenworthy, L., & Martin, A. (2011). Diminished sensitivity to sad facial expressions in high functioning autism spectrum disorders is associated with symptomatology and adaptive functioning. *Journal of Autism and Developmental Disorders*, *41*(11), 1475–1486. http://doi.org/10.1007/s10803-010-1170-0

Wallace, S., Coleman, M., & Bailey, A. (2008). An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition and Emotion*, *22*(7), 1353–1380. http://doi.org/10.1080/02699930701782153

Warton, D. I., & Hui, F. K. C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, *92*(1), 3–10.

Widen, S. C., Christy, A. M., Hewett, K., & Russell, J. A. (2011). Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition & Emotion*, *25*(5), 898–906. http://doi.org/10.1080/02699931.2010.508270

Widen, S. C., & Naab, P. (2012). Can an anger face also be scared? Malleability of facial expressions. *Emotion*, *12*(5), 919–925. http://doi.org/10.1037/a0026119

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research*, *102*(2), 109–124. http://doi.org/10.1016/S0165-1781(01)00225-6

Wilhelm, F. H., & Peyk, P. (2005). ANSLAB: Autonomic Nervous System Laboratory (Version 2.6). *Available at the SPR Software Repository: Http://www.sprweb.org*.

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*(4), 287–295. http://doi.org/10.1016/S0149-7634(01)00014-8

Wingenbach, T. S. H. (2010). *Feasibility and acceptability of a morphed faces emotion recognition paradigm - A pilot study*. University of Basel.

Winkielman, P., McIntosh, D. N., & Oberman, L. (2009). Embodied and disembodied emotion processing: Learning from and about typical and autistic individuals. *Emotion Review*, *1*(2), 178–190. http://doi.org/10.1177/1754073908100442

Wolpert, L. (2008). Depression in an evolutionary context. *Philosophy, Ethics, and Humanities in Medicine*, *3*(1), 8. http://doi.org/10.1186/1747-5341-3-8

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome Uusing the AQ: a preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. http://doi.org/10.1007/s10803-005-3300-7

Yoshikawa, S., & Sato, W. (2006). Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion. *Japanese Psychological Research*, *48*(3), 213–222. http://doi.org/10.1111/j.1468-5884.2006.00321.x

Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective & Behavioral Neuroscience*, *8*(1), 25–31. http://doi.org/10.3758/CABN.8.1.25

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests*. Bury St Edmunds, England: Thames Valley Test Company.